DOCUMENT RESUME

ED 071 778 PS 006 330

AUTHOR	Rentfrow, Robert K.
TITLE	Intensive Evaluation of Head Start Implementation in the Tucson Early Education Model.
INSTITUTION	Arizona Univ., Tucson. Arizona Center for Educational Research and Development.
SPONS AGENCY	Office of Child Development (DHEW), Washington, D.C.
PUB DATE	1 Aug 72
NOTE	58p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Affective Behavior; Analysis of Covariance; Attitude Tests; Behavior Rating Scales; *Child Development; Classroom Observation Techniques; Cognitive Development; *Comparative Analysis; Correlation; Data Analysis; Demography; *Disadvantaged Youth; Measurement Instruments; Preschool Education; *Preschool Programs; Program Effectiveness; *Program Evaluation; Social Development; Technical Reports; Test Reliability
IDENTIFIERS	Classroom Attitude Observation Schedule; Classroom Observation Procedure; McCarthy Scale of Childrens Abilities; Project Head Start; Schaeffer Behavior Inventory; Tucson Early Education Model

ABSTRACT

As part of the national Head Start Planned Variation Study, this study used a relatively small sample in an intensive evaluation of program implementation in one field community using the Tucson Early Education Model (TEEM). A modified Solomon four-group research design formed the organization framework. Evaluation of six TEEM classrooms and two locally implemented Comparison classrooms included a child data battery, classroom observation, situational tasks, Director's ratings, and demographic information. The child data battery, which emphasized both the cognitive and socio-affective domains, was given in the fall of 1971 to approximately half the TEEM children and to one of the Comparison classes. Spring testing included all groups. Analysis indicated that Comparison groups were not optimally matched to TEEM children; Comparison children were slightly older and performed better on pre-test. The difference in levels of inappropriate behavior during teacher absence between TEEM and Comparison classrooms was significant. Analysis of classroom observation indicated that TEEM classrooms had more child-initiated learning sequences combined with lower teacher initiated learning sequences. Due to limited sample size in this study, wide-ranging generalizations are not justified. But the utility of classroom observation techniques for assessing process goals within the open classroom framework was indicated. (Author/KM)

ED 071778

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Intensive Evaluation of Head Start Implementation in the Tucson Early Education Model

Robert K. Rentfrow
Principal Investigator

Kathleen Durning
Project Manager

Eva Conrad and Ocea Goldupp
Research Assistants

August 1, 1972

PS 006330

ARIZONA CENTER FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
College of Education
University of Arizona
Marsden B. Stokes
Director

FILMED FROM BEST AVAILABLE COPY

CONTENTS

FORWARD ..... inside front cover

ACKNOWLEDGMENTS ..... inside front cover

ABSTRACT iii

INTRODUCTION 1

Tucson Early Education Model 1

PAST RESEARCH 3

CURRENT RESEARCH DESIGN 5

RESULTS 13

Reliability 13
Pretest Data 14
Analysis of Covariance 15
Pre- to Posttest Gains 15
Solomon Four-Group Design 19
Posttest Data 24
Motivational Base 24
Classroom Process 30

DISCUSSION 43

Summary 45
Future Directions 45

APPENDICES 46

REFERENCES 54

ABSTRACT

As part of the national Head Start Planned Variation Study, this study used a relatively small sample in an intensive evaluation of program implementation in one field community using the Tucson Early Education Model (TEEM). A modified Solomon four-group research design formed the organizational framework for this study. Evaluation of six TEEM classrooms and two locally-implemented Comparison classrooms included a child data battery (selected subtests from the McCarthy Scale of Children's Abilities, and the Schaeffer Behavior Inventory), classroom observation (Classroom Observation Procedure), situational tasks (Classroom Attitude Observation Schedule), Director's ratings, and demographic information (Head Start Classroom Information Form).

Rather than relying on global test scores and emphasizing solely the cognitive development of the children, the child data battery emphasized both the cognitive and the socio-affective domains of development. The use of subtests allowed for the testing of program-specific goals.

The subtests used were from the McCarthy Scale of Children's Abilities, a recently developed test released this year by the Psychological Corporation. The Schaeffer Behavior Inventory enabled the classroom teachers to rate their pupils on three basic child classroom behaviors: Task Orientation, Extraversion and Hostility.

The Classroom Attitude Observation Schedule includes categories for classroom activities, grouping strategies, and the occurrence of inappropriate behavior in children. This instrument provides a picture of the children's self-direction by recording child behavior in the absence of a controlling adult.

The Classroom Observation Procedure, under development by Stanford Research Institute, includes a classroom check list, a verbal snapshot of the classroom grouping and activities, and a detailed recording of interactions. This allows the assessment of successful implementation of classroom processes within instructional models. Complete data analyses on this instrument were not possible due to the paucity of normative data presently available from SRI.

The child data battery was given in the Fall of 1971 to approximately half of the TEEM children, and to one of the two Comparison classes. Spring testing included the total TEEM group and both Comparison classes. Analysis indicated that the Comparison groups were not optimally matched to the TEEM children, and differences in both age and pre-treatment variables were evident with Comparison children being slightly older, and performing better on pretest. On 70% of selected subtests from the McCarthy Scale of Children's Abilities, the TEEM children achieved a greater raw score increase from pre- to posttesting than did the Comparison children. On two of these subtests, the difference in gains between the two groups was statistically significant ($p<.05$). A rank-order correlation between a rating of the teachers on their implementation of TEEM and their pupils' cognitive gains as measured by the McCarthy revealed the existence of a relationship between these two factors ($r = .63$ between implementation rating and child outcome).

The Classroom Attitude Observation Schedule data disclosed very little inappropriate behavior during teacher absent phase in the TEEM classrooms. The difference in levels of inappropriate behavior during teacher absent phase between TEEM classrooms and Comparison classrooms was statistically significant ($p<.05$). In the classroom rated as highest-implemented (from Director's ratings) the teacher participated with children in cognitively-oriented learning sequences. In a classroom rated as low-implemented, the teacher engaged in management activities, while children were mainly engaged in play-oriented learning sequences.

The Classroom Observation Procedure was utilized in the TEEM and Comparison classrooms to gather information on the nature of interactions between teachers and pupils. The variables were combined into eight variable constructs. The analysis indicated that TEEM classrooms had more child initiated learning sequences combined with lower teacher initiated learning sequences, when compared to the opposite pattern in Comparison classrooms.

Due to limited sample size in this study, wide-ranging generalizations are not justified. However, the utility of classroom observation techniques for assessing process goals within the open classroom framework was certainly indicated. Future research will be directed to the perfection of these techniques.

FORWARD

The Arizona Center for Educational Research and Development is a reorganized unit within the College of Education, University of Arizona, in Tucson. Historically, the major thrust of this group has been the development and implementation of new curriculum practices within the field of Early Childhood Education. Under the auspices of Head Start and Follow Through funds, a long-term program of research, development and field delivery has brought national attention to the University. However, newly emerging changes in national priorities within public education indicated that a broader base was needed for the diversification of research activities within the College of Education at the University of Arizona. The Arizona Center will continue its noteworthy work in the field of Early Childhood Education, while exploring new areas of research and development, both within the State of Arizona, and across the country.

ACKNOWLEDGEMENTS

The completion of a study of this magnitude would not have been possible without the able assistance of staff, teachers, and children. Special thanks are accorded to:

Mark Underwood
Becky Herndon
Ronald W. Henderson
Jean Godier
Shirley Hutcherson
Joseph M. Fillerup

Teachers, Aides and Children in the Head Start Program in Lincoln, Nebraska

This report was prepared pursuant to a contract with the Office of Child Development, Department of Health, Education and Welfare, Washington, D.C. The material herein is solely that of the authors, and does not represent official policy of either the Office of Child Development or the Arizona Center for Educational Research and Development. Any quotation or reproduction of material from this report should be appropriately cited.

INTRODUCTION

Preschool programs for disadvantaged children were introduced on a large scale with the passage of legislation under Title II of the Economic Opportunity Act of 1964. As part of a broad-based "war on poverty", one of the key provisions was the enhancement of early learning experiences of children from low-income families through a program called *Head Start*. Rather than proceed with a small-scale prototype phase, the Head Start program began nationwide in the summer of 1965, serving over 500,000 children in an eight week summer program. In retrospect, the know-how for accomplishing the broad-based goals of the Head Start program was quickly found lacking. This first summer program assuaged the national conscience, but many leaders in the fields of child welfare, psychology, and education were concerned about the reality of meeting the goals of the Head Start mandate. Since 1965, program efforts have been continuously modified to bring a focus on more intensive experiences with fewer children, and bring program documentation and evaluation questions to the forefront.

The study reported here is a result of that latter focus. As a part of the Planned Variation (PV) study of Head Start, the Tucson Early Education Model (TEEM) was selected as a program model for inclusion in a study of nationally disseminated early education programs. The study deals with questions relating to the development of program-specific evaluation tasks as applied to one field community using TEEM in Head Start. Evaluation studies of Head Start in the past have typically used large samples, and *post hoc* research designs (Cicirelli, 1969; Bissel, 1970). Results from these studies have been equivocal in showing lasting impact on young children's development. Thus, a need was seen for smaller studies which concentrated on program-specific research questions. This approach permits a more valid data base, and ultimately, conclusions with higher internal validity.

The goals of this study are:

1. To develop an observational-evaluation system to test selected process goals of the TEEM program.
2. To develop a program-valid set of tasks to assess children's development in Head Start classrooms using TEEM.
3. To field test this battery in a small study with six TEEM classrooms, and two locally implemented Comparison classrooms.
4. To perform statistical analyses of potential non-directional differences between these classrooms.

These goals were met through a combination of instrument development efforts at the Arizona Center, and a pre-and-post collection of data with children in Head Start classrooms in Lincoln, Nebraska. This research was initiated with the full involvement of the Head Start Director and staff in Lincoln. This kind of study would not have been possible without the full cooperation of the community and the children in the Head Start classrooms.

Tucson Early Education Model[1]

The Tucson Early Education Model was initially developed in 1965 as a cooperative project on the intellectual development of young Mexican-American children, conducted jointly by the College of Education, University of Arizona, and Tucson School District Number 1. Under the direction of Dr. Marie Hughes and Jewell Taylor, this project sought to identify causes of the high dropout rate of children from the Mexican-American community, and to develop a new educational program to enhance the educational experiences of these children. The continued development of this program became the focus of the Arizona Center for Early Childhood Education. In 1968, the Arizona Center was asked to become a sponsor for the Follow Through

[1]This section is adapted freely from a paper entitled "The Tucson Early Education Model", which was prepared as a program overview by the staff (especially Marie Hughes, Ralph A. Wetzel, and Ronald W. Henderson) of the Arizona Center for Early Childhood Education. A copy of the extended paper is available from the Information Officer, Arizona Center for Educational Research and Development, 1515 East First, Tucson, Arizona, 85719.

program, which was designed to provide high-quality early education programs for Head Start graduates. In 1969, the Office of Child Development asked the Arizona Center to design a complementary Head Start program in a number of the communities already using TEEM in Follow Through.

The rationale for TEEM is based on an understanding of the skills and abilities that are necessary to participate in contemporary America, as well as an appreciation of the varied backgrounds that children bring to the educational setting. The content and procedures of this program are, therefore, based on the definition and specification of the following:

1. The skills and attitudes necessary to function in our technical and changing society.
2. The behavioral characteristics which children bring to the educational situation.
3. The nature of the learning process.

The program procedures suggested by these considerations differ significantly from conventional curricula and modes of instruction for young children. If the requisite skills are to be developed, new program objectives and priorities must be established.

The major objectives of the Tucson Early Education Model can be classified into four categories.

Language Competence	Intellectual Base
Motivational Base	Societal Arts and Skills

Fig. I-1. Goal structure of TEEM.

1. **Language Competence.** Language competence is one of the major technical skills of the culture to which the children must adapt. Critical information is transmitted principally in verbal form. This requires an acquaintance with a variety of linguistic labels, concepts, language forms and an awareness of the function of language.
2. **Intellectual Base.** The intellectual base is a collection of skills assumed to be necessary in the process of learning. The skills are as yet only partially understood and defined, and are usually not formally taught. Yet their importance in every learning process is becoming increasingly recognized. Some of the intellectual base skills involve the organization of stimuli in the environment, e.g., ordering events along certain dimensions such as size, color and form. Some intellectual base skills are more complex behaviors which are difficult to define: to be able to attend, to recall significant events, to be able to organize one's behavior toward specific goals, to evaluate alternatives, and to choose, to plan and to develop expectations, to be able to discriminate significant and important behaviors in others and to imitate.
3. **Motivational Base.** By motivational base, we mean a collection of attitudes and behavioral characteristics related to productive social involvement. These include positive attitudes toward school and toward the learning process, an appreciation for learning and a willingness to persist at learning tasks, and an expectation of success and a willingness to change.
4. **Societal Arts and Skills.** Our culture is characterized by a wide range of arts and skills which constitute social interaction, information transmission, and scientific advance. Here we classify reading, writing, and mathematical skills as well as the social skills of cooperation, planning and democratic process. Although certain arts and skills have traditionally constituted the primary focus of school curricula, in the Tucson Early Education Model they are only a portion of the total program.

The purpose of the instructional program is to structure about the child a learning environment designed to promote the development of the behaviors defined by the four goal areas. Several aspects of the instructional program and process have been delineated.

1. **Individualization**. Based on the premise that children come to school with different sets of abilities and attitudes, then it is clear that teachers must individualize their teaching procedures. It is a characteristic of TEEM classrooms that frequent opportunities are provided for one-to-one adult-child interaction. A variety of behavioral options are constantly available to the child, providing opportunities to develop individual skills at individual rates.
2. **Imitation**. Although imitation is widely recognized as a significant process by which young children learn, it is seldom formally incorporated into classroom practice. The attention of children is directed toward the important and significant behaviors of others, and they are encouraged and reinforced for imitating. Imitation is a particularly important process in the acquisition of language. Adults work continuously to model elaborated and extended examples of the child's own language.
3. **Gratification**. Rewarding and gratifying experiences are clearly crucial elements in the learning process. Reinforcement plays an important role in classroom procedures. Every effort is made to ensure that the child experiences frequent gratification as a result of his behavior and skill acquisition.
4. **Generalization**. It is crucial to the success of an educational program that the skills which it teaches can be extended by the student to a variety of settings, objects and events. A skill is always taught in a functional setting, and is illustrated by a variety of examples in natural contexts.
5. **Orchestration**. The various skills which reflect the four goal areas are seldom exercised independently. It is a central aspect of TEEM that these skills are developed simultaneously through learning experiences that have interrelated goals. The technique of concurrently attending to and developing a variety of skills in children is defined as orchestration.

The major objectives of TEEM specify the development of skills and attitudes necessary to function in a technical and changing society. It is recognized, however, that in coping with their own environment, children develop other skills and attitudes which are quite functional and appropriate in their own homes and neighborhoods. Teachers and aides in the TEEM program learn to use the experiential backgrounds of pupils to further instructional objectives. The child's home and neighborhood are viewed as instructional resources, thus avoiding the discontinuity of cultural values which often confront minority children in school.

Implementation of the TEEM program in Head Start depends on a system of training and support services to communities. The delivery system depends on a multiplier effect, in which field representatives at the Arizona Center train program assistants in the communities, who then train teachers and aides at the classroom level. Pertinent training strategies and research findings, tested and demostrated at the local demonstration school in Tucson, are communicated through the delivery system to community staff for implementation in the classroom setting.

PAST RESEARCH

A major study of the long term effects of Head Start experience was conducted by Ohio University, in collaboration with the Westinghouse Corporation. This report, called *The Impact of Head Start* (Cicirelli, 1969), was a cross-sectional analysis of the performance of children who had a Head Start experience, and were in the early elementary grades (first through the third grade). Control groups were formed by matching each Head Start child with a child in his current class on selected demographic and educational variables. Children from around the country in the first, second and third grades were sampled in this study. This kind of design skirts the important task of specifying the classroom treatments in order to better understand the pattern of results. In the analysis, Head Start children were separated into two groups, one which had a full-year Head Start, and one which only had a summer experience.

The evaluation measures consisted of the Illinois Test of Psycholinguistic Ability, various measures of school readiness and achievement, and projective attitude measures completed by the children.

The overall conclusion was that no differences existed between Head Start children and control children on any of the above measures. There were differences between some subgroups, and across national regions, but this did not modify the overall null findings. The report concludes that summer programs appear ineffective in enhancing cognitive and affective growth. The report lacked any statement about the implementation of specific program models, and indicated wide variance among children and groups. It was thus concluded that Head Start programs were ineffective in influencing the development of disadvantaged children into the elementary grades.

The findings of the Westinghouse study were controversial, and many researchers criticized the design and analysis for being inadequate to the task (Madow, 1969; Smith & Bissell, 1970). This led to an intensive re-analysis of some of these data, in a study by Smith and Bissell (1970). They modified the analysis plan with new covariance procedures, and focused on first grade scores with urban black children. They concluded that the Head Start program was successful with this specific sub-group from the larger sample. They also noted that the most serious problem in the design of the Westinghouse study was the goal of assessing the "overall effectiveness" of Head Start through massive sampling of graduates. They called for, as others had before them, smaller studies which used quasi-experimental designs.

These criticisms of past research led to the specification of the longitudinal Head Start Planned Variation research design that is currently being conducted nationwide to assess the relative effectiveness of eight preschool program models on Head Start children, families and communities. This longitudinal study was modeled after the Follow Through program, and the Office of Child Development contracted with the Stanford Research Institute to collect data to evaluate all aspects of Head Start PV implementation. The first report on this study was produced in 1972, under the direction of Joan Bissell. In summarizing the mass of data collected on children, teachers and parents, it was noted that all Head Start programs, both model and "regular" classes produced gains in cognitive development and general achievement that were larger than could be expected by maturation alone. It was also noted that in the area of cognitive development, children in model programs made greater gains than children in regular classes. They note that, with two years left in the Planned Variation study, these findings are indeed preliminary, and await later replication with upcoming classes of Head Start children.

The above studies focused on large, nationwide samples of Head Start children, and a global set of measures. Another set of studies on Head Start age children have used small samples of children and intensive research designs to answer specific questions about preschool effectiveness. One of the most important research efforts in preschool program development has been the elaboration and comparison of competing curriculum models for use with disadvantaged preschoolers. The most extensive effort in this direction was a recently completed study which investigated the relative effects of three popular preschool models (Weikart, 1972). These three models were: 1) a **Language** model, modeled after the work of Becker-Englemann; 2) a **Cognitive** model, based on the theories of Jean Piaget; and 3) a **Unit-Based** or traditional nursery-school program. Curriculum variation was the critical research dimension, and other program components such as supervision and home contacts were held constant. The children entered at three years of age, and participated for two school years before entering public school kindergarten. All of the children were described as disadvantaged, and all met the criterion of being "functionally retarded" at their entry into preschool. Each classroom had two teachers, and one community aide, and all homes received a home visit once every fortnight. The research design included the investigation of the children's development of many measures of cognitive and socio-emotional growth, investigation of classroom teaching styles, and investigation of peer interaction processes.

Results of this three year project indicated that all children benefited from participation in preschool in terms of intellectual and emotional growth (Rentsow, 1971; Weikart, 1972). Differences in teaching style and peer interaction were minimal, and not significantly different. Weikart (1972) concludes that certain dimensions of program operation are necessary conditions for successful preschool intervention, independent of curriculum. The most important of these are teacher planning, and committment and provision of a growth-oriented model for supervision.

DiLorenzo and Salter (1968) reported a study of preschool programs incorporating eight different school systems assessing the longitudinal effectiveness of preschool for disadvantaged children. The children were selected as four year olds, and were randomly cast into treatment and

control groups. Each of the eight systems was free to develop its own model within the overall goals of the project. The specific curriculum foci ranged from Becker-Englemann to Montessori.

The evaluation design consisted of spring testing on standard measures of intellectual functioning (Stanford-Binet, Illinois Test of Psycholinguistic Abilities, Peabody Picture Vocabulary Test), and school readiness (Metropolitan Readiness). After the preschool year, the experimental groups made modest gains in IQ (1 to 4 points), as compared to slight losses for control children. Analyzing effects over all programs, they conclude that "The most effective prekindergarten programs were those with the most specific and structured cognitive activities" (p. 112). At the end of the kindergarten year, the experimental groups maintained superiority over the control groups in terms of school readiness. The data were analyzed by race and sex as well, and the non-white experimental children did not maintain their gains after kindergarten on measures of intellectual development. Also, the non-white experimental children were significantly lower than the white experimental children on the Metropolitan Test of school readiness.

CURRENT RESEARCH DESIGN

The previous review of research studies on Head Start, and Head Start-like, preschool programs provides an illustration of the dilemma that confronts evaluators in designing studies in this field. A research design must be judged both in terms of *internal validity* and *external validity* (Campbell & Stanley, 1963). The former relates to questions about the effectiveness of the treatment, in this case, implementation of a classroom program. The latter relates to the generalizability of the findings to other classrooms or other children. In retrospect, the Westinghouse study was designed to maximize external validity; that is, children from all regions of the country, from a variety of classrooms, were tested on global tasks. On the other hand, studies such as Weikart's comparative curriculum study concentrated on a small sample of children, randomly assigned to classroom models, and tested on more specific tasks relevant to program goals. This design seeks to maximize internal validity, at the expense of generalizability to programs and children all across the country. Campbell and Stanley treat these validities as interdependent; that is, internal validity is a precondition for external validity. These concerns have obviously influenced the Office of Child Development (OCD) in its current evaluation program for Planned Variation Head Start.

The national Planned Variation study was formulated to test the relative long-term efficacy of eight popular preschool models in Head Start. Cronbach (1963) described this kind of evaluation design as a "horse race" study. The data collected by Stanford Research Institute (SRI) as part of the national evaluation study of PV, on children, teachers and parents can be analyzed in terms of "win, place, and show" along different dimensions of Head Start goals. Concurrent with the SRI evaluation, OCD leadership decided to support small, sponsor-specific studies to help validate questions of program implementation and impact. These two kinds of studies then provide complementary information in terms of both internal and external validity.

Three different TEEM Follow Through sites were chosen for initial participation in the Head Start Planned Variation study. These communities were Lakewood, New Jersey; Walker County, Georgia; and Lincoln, Nebraska. These communities were selected in terms of having an effective and continuous feeder system of Head Start children into Follow Through classes of the same model sponsor. A fixed number of children in each community (about 120 each year) have participated in Planned Variation TEEM Head Start classrooms for the three year duration of the study. The national evaluation study conducted by SRI has collected information on children, teachers, and parents during these three years. These data are being interfaced with similar longitudinal information collected on Follow Through children. Ultimately, three cohorts of children will have experienced Head Start and Follow Through for five years under an integrated program model. These data will permit important questions to be answered concerning the relationship between program models and differential outcomes in children.

After the large national study of Planned Variation was initiated, the need for smaller, sponsor specific studies, was foreseen. Sponsors were asked in 1970 to propose discrete studies to provide valuable information on model-based questions. Of the three sites implementing TEEM Head Start, one was selected to provide the maximum amount of information within a limited budget.

The community selected for this study was Lincoln, Nebraska. This middle-sized community in the Great Plains has maintained a strong, community-based commitment to the Tucson Early Education Model since its introduction in 1969. The Head Start Director has been involved in the development of programs for low-income children in Lincoln since their inception. The community's commitment to TEEM has been seen in wide participation in TEEM training sessions by staff from the larger school system. The program has also been noted by an active and well accepted parent involvement component. The community is predominantly Anglo, with a small percentage of Negro children (10% in this sample), and virtually no other ethnic groups.[2] The predominant family background of Head Start children, then, is of families that have immigrated from the small farms of the Plains and Southeast. These families are distributed widely across the metropolitan area, and there is no core, or "ghetto" area of poverty. In this community, during the 1971-72 school year, Head Start children were placed in classroom locations within elementary schools in the community. Thus, most of the children were bused to their Head Start classrooms.

The selection of Lincoln for the study was premised on a high-implementation site that offered full cooperation to the program sponsor in doing evaluation research. Stable implementation is assumed to be a necessary (but not sufficient) condition for influencing children's development. Medley and Mitzel (1963) note that our concern for the implementation of educational systems must ultimately be focused in terms of effects on pupils, or changes in pupil behavior. With this premise, the following study was designed to specify the goal areas of TEEM Head Start in terms of operational definitions, and to use these domains in comparing the growth of children in six TEEM implemented Head Start classes with children in two locally implemented classes. Another element in the design is the development and validation of a new technique to assess process outcomes in TEEM classrooms.

The selection of dependent variable measures was based on the following considerations. The TEEM model puts an emphasis on children's development in both the cognitive and the socio-affective areas. Traditionally, evaluation designs have leaned heavily on the former as the major outcome area. The design outlined for this study puts an equal emphasis on both domains. Rather than depend upon global scores, such as IQ, this design depends upon the specification of a set of tasks which have face validity in the TEEM program to evaluate a specified set of outcomes. Also, the design is explicitly child-based, since a wide data base concerning teacher training and implementation is emanating from the TEEM Follow Through program.

The selected dependent variable measures, grouped by goal areas, are given below.

Intellectual Skills
- Pictorial Memory – MSCA[3]
- Verbal Memory
- Conceptual Grouping

Motivational Base
- Schaeffer Behavior Inventory

Societal Arts and Skills
- Counting and Sorting
- Imitative Acting
- Drawing
- Number Questions

Language Competence
- Word Knowledge
- Verbal Fluency
- Opposite Analogies

[2]These comments are based on information gathered on the *Head Start Classroom Information Form*, which summarized information on various family attributes, such as income, parent's education, parent's place of birth, and vocational history.

[3]All of the subtests, except the Schaefer Behavior Inventory, are from the McCarthy Scale of Children's Abilities. This test was utilized in pre-published form through arrangements made with Dr. Alan Kaufman, Research Psychologist at the Psychological Corporation. Staff at the Arizona Center had been involved in the early standardization of the McCarthy, and felt it was a promising new direction for assessing development in young children. Psychological Corporation expects to release the test publicly by September, 1972 (personal communication, June, 1972).

The child-based data set utilizes subtests from the McCarthy Scale of Children's Abilities. This test, recently developed by the Psychological Corporation, represents the culmination of the work of Dorothea McCarthy, a prominent child psychologist. The Scale utilizes both paper and pencil and toy-like materials in the individually-administered assessment of intellectual and behavioral development of children 2½ to 8½ years of age. The goals which guided test development were to make the subtests (a) intrinsically interesting to young children, (b) emotionally neutral, and (c) relatively quick to administer in order to permit the sampling of a wide variety of behaviors and avoid taxing the short attention span of children. The selection of subtests for this study was based on the need to explicate specific performance domains rather than depend on total test scores. The number of items per subtest varies, but generally a pattern of increasingly difficult items is followed. Clustered by the TEEM-specified goal areas, the subtests employed in this study are further explicated in Figure I-2.

Children from culturally different backgrounds often face the educational environment with a negative perception of self-worth. One of the intended impact areas of TEEM is to improve this self-concept as a facilitating condition to other learning experiences. The *Preschool Self-Concept Picture Test* (PSPT) was developed by Wollner for use with preschool children in Head Start sites to assess the attitudes that children have toward themselves. The test format contains equivalent picture stimuli for both Caucasian and Negro children, and forms for boys and girls. The picture stimuli represent ten criterion attitudinal dimensions, presented in bi-polar fashion (e.g., dirty-clean; independent-dependent). The child selects the one which he feels represents himself. This technique yields a self-positivity score, which may be analyzed in a pre-post design.

The Preschool Self-Concept Picture Test, initially intended for use in assessment of motivational base, was discarded for further use after pretest analysis. Scored for positivity of self-concept, the uncorrected split-half reliability estimate was -.065. Choices scored as "positive" such as "strong" for boys and "weak" for girls lack even face validity in many cases, and the null split-half consistency reveals that there is no construct, such as positivity of self-concept, underlying the items. It is doubtful whether four-year-olds are defining self-view in terms of the relevant attributes when asked to choose between two pictures by the phrase "Which boy are you?" This is further revealed by the fact that when asked "Which boy would you like to be?" (ideal self-concept) the children chose the so-called "negative" picture more frequently than when defining actual self-concept on over half of the attributes (e.g., 67% of the children defined themselves as happy by initial picture choice, but only 61% supposedly wished to be happy when making the picture choice for ideal self-concept).

It was decided to substitute the Schaeffer Behavior Inventory (SBI) for the PSPT to assess motivational base growth in children. The SBI, as adapted by Stanford Research Institute, is a fifteen-item checklist designed to assist teachers in assigning scores to three basic child classroom behaviors. These three behaviors are:

1. Task Orientation – how well a child attends to and stays with classroom activities.
2. Extraversion – how readily a child interacts with other people.
3. Hostility – how a child responds to some of the adjustments and conflict problems encountered in group activities.

The scale used for each item is of the Likert-type, with seven options: never, almost never, occasionally, half the time, frequently, almost always, always. For data analysis, a value of one is assigned to "never" and a value of seven to "always". Under the auspices of the Stanford Research Institute study, the six TEEM Head Start teachers completed an inventory on each of their students in the early Fall of 1971. Taking advantage of this available pretest data, TEEM teachers were asked during posttesting to re-rate children in each of these classes. Teachers in the Comparison classrooms rated their children in April only.

The *Head Start Classroom Information Form* (HSCIF) was developed by Stanford Research Institute to gather summary information on background variables considered important in understanding the development of children in low-income families. Many studies have indicated that home environment variables are important predictors or covariates in understanding the relationship between children's learning and their school context (Hess & Shipman, 1965; Henderson, 1972). The HSCIF asks questions about family income level, occupational histories of the parents, educational backgrounds of the parents, and the presence of amenities in the home. The form is completed by teachers from the Head Start classes, and typically requires home contacts with family members. These data were collected in Head Start classes during the

Subtests	Stimulus Materials	Child's Task	Timed
INTELLECTUAL SKILLS			
Pictorial Memory	Colored cards, picturing 6 common objects per page (e.g., penny, spoon, pin, cowboy, key, watch).	Is to name as many of the objects as he can remember after viewing the card for 10 seconds and the card is removed from sight.	Yes
Verbal Memory	Verbal presentation of sequence of 3 and 4 words, 14-word sentences and a one-paragraph story.	Is to repeat as much of the stimuli as he can remember.	No
Conceptual Grouping	Set of 12 blocks: 6 squares, 6 circles, each shape in 3 colors (red, yellow, blue) and 2 sizes per color.	Is to perform various manipulations along the 3 dimensions of color, size and shape according to verbally presented instructions (e.g., item 5: "Find all the big yellow blocks.").	No
Numerical Memory (given on the posttest only)	Verbal presentation of digit spans from 1-6 digits in length.	Is to repeat the stimulus digits either forward or backward in accordance with the instructions.	No
SOCIETAL ARTS AND SKILLS			
Counting and Sorting	10 1-in. green cubes; 2 cards picturing various arrangements of dots.	Is to demonstrate knowledge of ordinal and cardinal numbers and such number-relevant vocabulary as all, same, some (e.g., Part I, item 8: "Show me the second block from this end.").	No
Imitative Action	The tester models various simple behaviors, (e.g., crossing feet, folding hands, twiddling thumbs, looking through tube).	Is to mimic. Full credit is earned for exact mimicry (e.g., putting the right thumb over the left as modeled).	No
Drawing	Pieces of unlined paper and pencil. Part 1: the tester models various geometric designs; Part II: the tester instructs the S to draw a human figure of the same sex as the S.	Is to follow the modeled drawings and instructions.	No
Number Questions	Verbally presented questions and problems.	Is to respond with numerical answers (e.g., item 12: "If I went to the store and bought a dozen apples, how many apples would that be?").	No
LANGUAGE COMPETENCE			
Word Knowledge	Part I: Picture cards similar to those used in Pictorial Memory; Part II: the verbal presentation of vocabulary words.	Is to recognize common items presented on the cards, to name common items and to define the words orally presented. (e.g., coat, thread, month).	No
Verbal Fluency	Verbal request for subordinate members of classes.	Is to name as many subordinate members as possible within 20 seconds (e.g., item 2: "Lets see how many different kinds of animals you can think of before I say stop.").	Yes
Opposite Analogies	Verbal presentation of a simple analogies.	Is to provide the missing word in the analogy (e.g., item 5: "Cotton is soft, and rocks are _____.").	No

Fig. I-2. Selected subtests from MSCA.

Fall of 1971. Much of the same information on the Comparison classes was collected during the Spring of 1972. Since these are essentially stable variables, no bias is expected between these groups as a result of different collection times. For the Comparison group, only parent education, family income and family occupation were collected. The description of the Lincoln Head Start families, based on these data, was described earlier (see page 6).

Another major goal of this study was the development of a new technique for assessing certain "process goals" within the TEEM Head Start classroom. These process goals have consistently eluded formal evaluation, since young children don't respond meaningfully to paper-and-pencil attitude or self-concept tests. Working from the Classroom Observation Procedure developed by SRI, combined with procedures developed at the Arizona Center, a new technique incorporating classroom observation techniques with environmental manipulation was developed.

The Classroom Attitude Observation System (CAOS) includes categories for classroom activities, grouping strategies, and the occurrence of inappropriate behavior in children. The observation is divided into three phases: *baseline* phase of 12 minutes; *teacher absent* phase of 12 minutes, in which teacher and aides are excused from the room; and a *reinstitution* phase of 12 minutes, with teacher figures again present. It is hypothesized that children in a TEEM classroom develop internalized self-direction, and do not need the external, implied control of teacher presence to continue patterns of learning behavior.

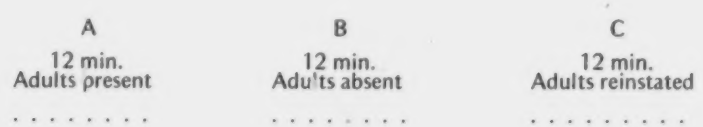


Fig. I-3. Phases of CAOS.

The observational technique counts children and adults engaged in the various classroom activities on a time-sample basis. All of the types of activities presumed to take place in the classroom are listed on the recording form (Figure I-4). Once every two minutes a clockwise visual scan is made of the room by the observer. The observer remains stationary throughout the thirty-six minute period, as the scan begins and ends at the same point for each scan. Numbers of children and adults observed during that scan are placed in the appropriate cell, while retaining grouping patterns in the recording. If inappropriate behavior is observed during this scan, it is also noted by its associated activity and in the appropriate two-minute scan period. Two more scans are made during the two-minute period to pick up incidents of inappropriate behavior, once at the end of the minute, and again at the end of a minute and a half.

Procedures and categories from two previously developed instruments were synthesized to produce this particular procedure. Both of the parent procedures have been field tested and found reliable.

One of the two instruments was developed by Dr. Jane Stallings, of Stanford Research Institute. It is an interaction observation technique developed for National Head Start and Follow Through evaluation efforts, called Classroom Observation Procedure. As a preamble to each five-minute interaction recording period, the observer takes a "snapshot" which gives the foundation for the CAOS technique. Activity categories are taken directly from the "snapshot". Within SRI's coding system, grouping patterns are classified as "i" for "individual", "2" for "pairs", "S" for "small groups" (three to eight children), and "L" for "large groups" (more than eight children). Given this system the total number of children engaged in an activity at a given time may not be apparent. With CAOS, numbers are used for all groups recorded in order to add this dimension to the procedure. Activity definitions used by SRI are more encompassing than those used for CAOS, in order to include all types of curricula and all grade levels involved in Head Start and Follow Through. While activity definitions for CAOS follow SRI's system essentially, they have been redefined to make them more program specific to the Tucson Early Education Model, and to deal more appropriately with pursuits of four and five year old children (Goldupp, 1972). Figure I-5 is an example of such redefinition.

CLASSROOM ATTITUDE OBSERVATION SCHEDULE (CAOS)

page ___ of 3

Teacher _____ School _____ Community _____ Date _____ Observer _____ Time Started _____

	N.I.	I.A.	N.I.	I.A.	N.I.	I.A	N.I.	I.A.	N.I.	I.A.	N.I.	I.A.	Total
A. Snack Lunch													
B. Group time													
Story/Sing Dance/Music													
C. Arith/Math Numbers													
Reading/Alpha Lang. Devel.													
D. Soc. Studies Geography													
Science Nat. World													
E. Games Puzzles													
F. Arts Crafts													
Sew/Cook Pound/Saw													
G. Blocks Trucks													
Dolls/Dressup Playhouse													
H. Play													
I. Transitional activities													
J. Classroom management													
K. Out of room													
L. Observing Other													
M. Wandering													
Totals													

N.I. = Number Involved
I.A. = Inappropriate Activity

Revised 8/72 Development Copyright, Arizona Center for Educational Research and Development, Ocea Goldupp & Robert K. Rentfrow, Ph.D.

Fig. I-4. CAOS form.

SRI Version	CAOS Version
"This category refers to the teaching and learning about plants, animals, minerals (care of, collection, comparison) and science concepts. It is subdivided into texts, workbooks; plants, animals; science equipment; and films, slides as tools for teaching and learning in this area. All of these have been defined except for science equipment, which refers to any apparatus or concrete objects used in the course of teaching and learning about science and the natural world."	This category refers to teaching and learning about plants, animals, minerals (care of, collection, comparison) and science concepts. With very young children this may consist of: playing with a classroom animal such as a rabbit or gerbil looking at fish in a tank looking at objects through a microscope. Included in the science concepts would be the intellectual skills taught through a variety of means. This would include "Intellectual kits" which are used for fostering observational skills, and for making comparisons, among other things. This will also include activities dealing with shapes and the discrimination skills associated.

Fig. I-5. Comparison of SRI and CAOS definition for one activity category: "Science, Natural World".

SRI found the snapshot portion of their procedures highly reliable (better than 90% over thirty trainees in a stationary test situation) and the training of observers relatively simple.

The other parent instrument was developed at the Arizona Center for Early Childhood Education by Drs. Sadie Grimmett and Billie Underwood. The Schedule for Incompatible Learning Behavior (SILB) (Grimmett, Underwood & Brackney, 1970) was developed for a study assessing the relationship of behavior settings to disruptive or inappropriate behavior. With some modification, the CAOS categories and definitions for inappropriate behavior were borrowed from this instrument. The categories consist of "hitting", "yelling", and "other" to replace SILB's "disturbing". If a behavior was observed that was clearly disruptive but not part of the existing categories, the observer coded this "other" and noted the behavior at the bottom of the observation form.

The population of behavior sampled was that occurring during *free choice time* in all classrooms. This is the time when children choose their own activities from those available in the classroom. The decision to standardize procedures on free choice time stemmed from a combination of reasons:

a) The behavioral setting needed to be consistent across all classrooms. Grimmett et al. (1970) found powerful evidence that the behavioral setting controls incidence of disruptive behavior.
b) Individual choice time was associated with lower rates of inappropriate behavior than large group time (Grimmett et al., 1970).
c) The block of time during which children choose their own activities was the longest time segment in both TEEM and locally implemented Head Start Classrooms.

Observations took place the last week of April, 1972, with one thirty-six minute observation period for each of eight classrooms during that week. Two observers observed in four classrooms each. Training consisted of a review of categories and their definitions, and practice coding by both observers in the same classroom, without the manipulative phase. After practice coding, the two observers met to compare coding and isolate trouble spots. When definitions were not clear, they were re-worked until both observers could agree on their meaning and observability. The observers achieved reliability, with average agreement of 82%, and a high of 91% (Scott's pi; c.f. Flanders, 1966). It was felt that this was high enough to permit comparability of observations made by the observers in separate observations.

The Classroom Observation Procedure (COP), developed and fielded by staff from Stanford Research Institute, provided additional information on the operation of the classrooms within this study. Staff from the Arizona Center visited SRI in Palo Alto, California, in April, 1972, and coded all of the COP protocols taken on the TEEM classrooms during the Spring of 1972. The Arizona Center contracted directly with the SRI trained observer to observe in the additional Comparison classrooms in Lincoln. The COP collects information on the physical plant, type of ongoing classroom activity and locations of teachers and learners, as well as ongoing-teacher-child interactions. Twenty-four five-minute observation periods comprise the information on each classroom. The mass of data was reduced through computer programs developed especially for the COP system by programming staff at the Arizona Center.

The Head Start Director in Lincoln was asked by SRI as part of their national study to rate all of the TEEM classrooms in Lincoln. This rating was done on a nine-item Likert format, with overall implementation being the variable under scrutiny. The Arizona Center asked the Director to add the two Comparison classrooms into this format, and rate all of the classrooms on the same form. These data comprised the independent ratings of TEEM implementation.

The design of the study was based on a Solomon 4-group paradigm, which introduces certain economies in costs, as well as providing answers to questions about the potential artifact of repeat testing on group means. The Solomon design involves pretesting only 50% of the treatment and control groups (subjects are randomly selected).[4] Posttest is administered to all subjects in both groups. This ideal design was compromised in the following ways. At pretest time, one total class of Comparison was randomly selected, rather than selecting subjects randomly. Additionally, due to high attrition in children this age, all subjects were posttested, whether they had been pretested or not. See Figure I-6 for graphic design of this study.

[4]The N's given in Fig. I-6 are based on the proposed operation for the Head Start program in Lincoln in 1971-72. Due to attrition and absences, the N used in the analyses is slightly smaller.

Child Data Battery

McCarthy Scale of Children Abilities

- Word Knowledge
- Counting and Sorting
- Verbal Memory
- Drawing
- Verbal Fluency
- Opposite Analogies
- Conceptual Grouping
- Pictorial Memory
- Imitative Action
- Number Questions

Schaeffer Behavior Inventory (SRI)

Classroom Observation

Classroom Observation Procedure (SRI)

Situational Tasks

Classroom Attitude Observation Schedule

Demographic Information

Head Start Classroom Information Form

TEEM Implementation Rating

Director's Rating

	N	Pretest Sept., 1971	November	February	N	Posttest April, 1972
TEEM Head Start	N~60	Child Data	Demographic Data	Class Observation	N~120	Child Data
Comparison	N~12				N ~24	Director's Rating Situational Tasks

Fig. I-6. Selected instrumentation.

RESULTS

In the following discussion, there are three distinct data points that will be reported. The first analysis is concerned with pretest scores only; the second is concerned with pre-to-post gain scores for children who remained the whole year; the third is a post-only analysis for all children in classes at the end of the year. The second group, subjects who were tested at both data points, provides the most important data for analysis. The reader is cautioned to follow closely the different levels of analysis as pointed out above, since the power of certain analyses is a function of which sample is being reported.

Reliability

Sixty-four of the children tested in the Fall of 1971 were also available for testing the following Spring. Test-retest reliability coefficients on the McCarthy subtests and a McCarthy summary score were calculated on this sample (Table II-1). The McCarthy summary score is a partial total, since it is based on the raw score sum of only the ten subtests which was selected for use in the TEEM evaluation from the complete McCarthy Scale of Children's Abilities. Reliability estimates on the subtests ranged from .18 to .78, with a mean pre-post correlation of .55 across all ten. The McCarthy summary score displayed a test-retest reliability of .87. This represents a considerable degree of stability given the seven-month interval between administrations, and any effects of intervening treatments. The very low pre-post correlation for the Imitative Action subtest may be explained by the presence of a ceiling effect. Near-perfect and perfect scores were frequent. On the pretest, children averaged 4.02 points out of a maximum possible of 5.00, while the posttest mean was 4.34.

Table II-1 also reveals subtest-total correlations between each subtest and the pretest McCarthy summary score. These correlations have been corrected for the contribution of the subtest score to the summary score. The subtest-total correlations range from a low of .05 for Imitative Action

TABLE II-1

McCARTHY TEST-RETEST RELIABILITIES AND SUBTEST-TOTAL CORRELATIONS

	Test-retest correlations	Corrected subtest-total correlations
Pictorial Memory	.34	.32
Verbal Memory	.61	.61
Conceptual Grouping	.78	.70
Counting & Sorting	.62	.72
Imitative Action	.18	.05
Drawing	.75	.64
Number Questions	.42	.39
Word Knowledge	.57	.69
Verbal Fluency	.66	.66
Opposite Analogies	.61	.57
McCarthy Summary Score	.87	

Note.—Reliabilities and subtest-total correlation are based on 64 children who received both pre- and post-McCarthy (52 TEEM and 12 Comparison children).

to .72 for Counting and Sorting, with an average subtest-total correlation of .54. The correlation involving the former subtest is again hampered by ceiling effects, but it is not unexpected that Imitative Action would show little commonality with the generally more cognitive abilities tapped by the other McCarthy subtests. The subtest-total correlations, with the exception of that for Imitative Action, reveal a fair degree of common variance among the subtests despite the diversity of abilities tapped.

Pretest Data

Seventy-seven children were individually tested on selected McCarthy subtests in the fourth week of September, 1971. Those tested represented approximately half (N=62) the TEEM children, randomly selected from each of the six programs classrooms. In addition, one of the two Comparison classes were tested in its entirety (N=15). A second Comparison classroom became available for testing at a later date, and was sampled at posttest only.

Data analysis revealed that the Comparison children scored consistently higher across all McCarthy subtests on the pretest than did TEEM children (Table II-2). Four of these 10 mean differences achieve statistical significance in favor of Comparisons: Drawing, Number Questions, Word Knowledge, and Verbal Fluency.

TABLE II-2

McCARTHY PRETEST MEANS AND F-RATIOS FOR TEEM AND COMPARISON HEAD START CHILDREN

	TEEM N=62	Comparison N=15	Test of Difference
	$\bar{X}$ Raw Score	$\bar{X}$ Raw Score	F-Ratio
INTELLECTUAL SKILLS			
Pictorial Memory	4.43	5.67	3.72
Verbal Memory	17.89	21.93	1.94
Conceptual Grouping	5.63	6.20	1.24
SOCIETAL ARTS AND SKILLS			
Counting and Sorting	5.58	6.27	1.07
Imitative Action	4.10	3.87	3.15
Drawing	6.39	8.80	4.84*
Number Questions	3.48	4.07	5.14*
LANGUAGE COMPETENCE			
Word Knowledge	13.58	16.13	4.87*
Verbal Fluency	6.52	9.20	4.32*
Opposite Analogies	3.32	3.80	1.25

*$p<.05$.

As is often the case in quasi-experimental designs, the Comparison group was clearly not drawn from a population comparable to the TEEM children. The striking dissimilarity in pretest performance existed before any differential treatments were instituted.

Analysis of family demographic variables revealed no statistically significant socioeconomic differences between TEEM children and the more able Comparison pupils (Appendix A), though

Comparison children were slightly higher in mean family income and rate of employment and education of head of household. No doubt more important in accounting for the superior pretest preformance of Comparison students was the fact that Comparisons were, on the average, six weeks older than TEEM children.

Several solutions to the problem of the noncomparable control group were considered. Matching of Comparison children to selected experimental subjects on the basis of pretest scores was not feasible due to the lack of overlap between the two pretest distributions. Of the 64 pretested subjects who were also available at the time of the posttest, 24 of the 52 TEEM children and only one of the 12 Comparisons (46 and 8% of the two groups, respectively) scored below 67 (McCarthy summary score, pretest). Conversely, in the upper ranges, four Comparisons (33%) but only one TEEM subject (2%) scored above 105.

Analysis of covariance and gain score procedures were considered as post hoc methods of statistically equating the groups.

Analysis of Covariance

Of the 12 Comparison and 52 TEEM pupils who received both pre- and post- McCarthy tests, the Comparisons were significantly higher than TEEM subjects on the pretest summary score (means of 89.17 and 67 84 respectively; $F=8.24$, $p<.01$). Although the difference in summary totals between the two groups was reduced, Comparisons maintained their superiority on posttesting (means of 116.08 and 98.56, $F=4.45$, $p<.05$).

Analysis of covariance with pretest as the covariate and posttest the dependent variable in a situation matrix where Comparisons have higher pretest scores results in undercorrection for pretest differences, and makes the experimental treatment look damaging, even in a simulated case where there are actually no treatment effects (Campbell and Erlebacher, 1970, pp. 196-197). It is therefore especially noteworthy that such a covariance procedure on the present data resulted in a *higher* adjusted posttest mean on McCarthy summary score for TEEM children than Comparisons, equating on pretest, even though the difference was not significant (102.43 and 99.33, respectively; cf. Appendix B).

Because of the limitations of covariance and the impossibility of matching, the analysis of program effects was carried out in terms of gain scores, the least questionable procedure in view of the present circumstances.

Pre- to Posttest Gains

Program effectiveness may be assessed in terms of the amount of improvement attained by the children between the two testing sessions. Further, such an analysis in terms of gain scores allows comparisons between control and experimental groups, despite initial pretest differences.

On all subtests except Imitative Action, there were statistically significant gains made by the TEEM children from the pre- to posttesting (Table II-3). On most subtests the Comparisons also made significant gains (Table II-4), though the small sample size for Comparisons prevented several sizeable increases (e.g., Verbal Memory) from achieving statistical significance.

Table II-5 indicates that on 70% of the subtests the TEEM children achieved a greater raw score increase from pre- to posttesting than did the Comparison children. These subtests are Pictorial Memory, Verbal Memory, Conceptual Grouping, Counting and Sorting, Number Questions, Word Knowledge and Opposite Analogies. On two of the seven subtests (Pictorial Memory and Word Knowledge) the differences in gain between the two groups are statistically significant at $p<.05$. The tendency for greater absolute gains by TEEM children persists despite the fact that treatment effects relative to control group gains are usually underestimated by gain score procedures when controls show initial pretest superiority (Campbell and Erlebacher, 1970, pp. 197-198).

Figures II-1 through II-4 present the pretest (Fall) and posttest (Spring) raw scores on the selected McCarthy subtests for both the Comparison and the TEEM children. The blackened sections of the graphs elucidate the gain scores, or the improvement in subtests scores from the pre- to the posttest. In only one instance (Figure II-2) is a decrease in scores seen from the Fall to the Spring testing. The maximum possible score is indicated beneath each subtest title. The McCarthy Scale of Children's Abilities is designed for children from age 2½ to age 8½. Hence the appropriate normative score is a more valid criterion than is the maximum possible score. On all graphs the dashed lines present the normative scores for five year old children. This is somewhat misleading, however, since there is a difference in the group mean ages, with the Comparison

TABLE II-3

PRE- AND POST-McCARTHY SUBTEST MEANS AND F-RATIOS
FOR TEEM HEAD START CHILDREN

	Pretest N=52	Posttest N=52	Gain	Test of Gain
	$\bar{X}$ Raw Score	$\bar{X}$ Raw Score		F-ratio
INTELLECTUAL SKILLS				
Pictorial Memory	4.40	5.85	+ 1.45	15.91**
Verbal Memory	16.52	22.37	+ 5.85	24.51**
Conceptual Grouping	5.37	7.50	+ 2.13	56.99**
SOCIETAL ARTS AND SKILLS				
Counting and Sorting	5.29	8.06	+ 2.77	66.63**
Imitative Action	4.04	4.29	+ 0.25	1.93
Drawing	6.02	14.85	+ 8.83	168.89**
Number Questions	3.67	4.96	+ 1.29	41.07**
LANGUAGE COMPETENCE				
Word Knowledge	13.25	16.67	+ 3.42	45.79**
Verbal Fluency	6.15	9.67	+ 3.52	48.95**
Opposite Analogies	3.13	4.35	+ 1.22	49.48**

Note.—Statistically significant gains are evidenced by significant trials main effects (pretest to posttest) in analyses of variance.
**$p<.01$.

TABLE II-4

PRE- AND POST-McCARTHY SUBTEST MEANS AND F-RATIOS
FOR COMPARISON HEAD START CHILDREN

	Pretest N=12	Posttest N=12	Gain	Test of Gain
	$\bar{X}$ Raw Score	$\bar{X}$ Raw Score		F-ratio
INTELLECTUAL SKILLS				
Pictorial Memory	6.08	5.75	− 0.33	<1
Verbal Memory	22.58	27.24	+ 4.66	4.17
Conceptual Grouping	6.92	8.33	+ 1.41	6.48*
SOCIETAL ARTS AND SKILLS				
Counting and Sorting	6.58	8.83	+ 2.25	10.40**
Imitative Action	3.92	4.58	+ 0.66	6.77*
Drawing	10.08	20.58	+10.50	50.70**
Number Questions	4.00	4.67	+ 0.67	3.52
LANGUAGE COMPETENCE				
Word Knowledge	16.75	17.83	+ 1.08	<1
Verbal Fluency	8.33	13.67	+ 5.34	30.12**
Opposite Analogies	3.92	4.58	+ 0.66	2.84

Note.—Statistically significant gains are evidenced by significant trials main effects (pretest to posttest) in analyses of variance.

*$p<.05$.
**$p<.01$.

TABLE II-5

RAW SCORE GAINS ON McCARTHY SUBTESTS FROM FALL TO SPRING FOR COMPARISON AND TEEM CHILDREN

	Mean Gains		Test of Difference in Gains
	Comparison N=12	TEEM N=52	F-ratio
INTELLECTUAL SKILLS			
Pictorial Memory	− 0.33	+1.45	4.89*
Verbal Memory	+ 4.66	+5.85	<1
Conceptual Grouping	+ 1.41	+2.13	1.23
SOCIETAL ARTS AND SKILLS			
Counting and Sorting	+ 2.25	+2.77	<1
Imitative Action	+ 0.66	+0.25	1.11
Drawing	+10.50	+8.83	1.12
Number Questions	+ 0.67	+1.29	2.88
LANGUAGE COMPETENCE			
Word Knowledge	+ 1.08	+3.42	3.89*
Verbal Fluency	+ 5.34	+3.52	2.50
Opposite Analogies	+ 0.66	+1.22	1.81

Note.—Statistically significant differences between gains of TEEM and Comparison children are evidenced by significant groups (Comparison-Experimental) by trials (pre- to post-McCarthy) interactions in the repeated measures analyses of variance.

*$p<.05$.

group being six weeks older than the TEEM children. Appendix C reveals that a dissimilarity in age at the time of the posttest is particularly worthy of notice. Normatively, there is a greater increase in raw scores for the six month period between 4½ and 5 than 4 and 4½. The six week age variance is, hence, more crucial when analyzing the posttest than the pretest data.

Figure II-1 presents the pretest and gain scores attained on the three general goal areas of *Intellectual Skills, Societal Arts and Skills,* and *Language Competence.* These raw score figures were determined by the addition of the subtest raw scores subsumed under each of these three general variables. Although the Comparison subjects began the school year with a higher mean raw score and ended the year closer to or above the norm, the TEEM children gained more in Intellectual Skills and in Language Competence, by 3.6, and 1.1 points respectively. On the Societal Arts and Skills cluster the Comparison children gained seven-tenths (0.7) of a point more than the TEEM children.

Figure II-2 illustrates the raw scores for subtests included in the general category of Intellectual Skills. Refer to Figure I-2 for a detailed account of the content of the various subtests.

The Comparison group decreased from the pre- to the posttesting on the Pictorial Memory subtest. Consequently, their posttest score is below that of the TEEM children although their pretest score is higher. On all of the subtests included in the Intellectual Skills variable the Comparison children's pretest score was higher but the TEEM children gained more from the pre- to the posttesting. Neither group achieved the normative score on any of the subtests.

The data for the subtests which comprise the evaluation of Societal Arts and Skills are shown in Figure II-8. On all four of these subtests, except Imitative Action, the Comparison children's pretest score was higher than the TEEM children's. Gains were greater for the latter group on Counting and Sorting and on Number Questions, and TEEM students more than doubled their pretest score on Drawing. The actual numerical gain by the Comparison group on this subtest was greater, as was their gain on Imitative Action. The Comparison group met the norms for five-year-olds on the Drawing subtest.

The Language Competence variable included three subtests, and these scores are presented in Figure II-4. On all of these subtests the Comparison children began at a higher level than TEEM children, and on two of these three the TEEM group made higher gains. The Comparison group met the normative mean on the Verbal Fluency subtest, and both groups were within one half (0.5) of a point of meeting the norm on the Opposite Analogies subtest.

Analyses were run to determine if there were differential gains from pre- to post-McCarthy by sex of the children. No such interactions were evident in terms of McCarthy summary score or the three TEEM goal areas of Intellectual Skills, Societal Arts and Skills, or Language Competence. One statistically significant interaction between sex and gains occurred on the Conceptual Grouping subtest ($F=10.97, p<.01$). This may be explained by the fact that independent of the comparison-experimental distinction, females (N=29) gained an average of 2.9 points on the subtest pre- to post, while males (N=35) gained only 1.3 raw score points.

The coordinator of Head Start programs for the Lincoln area rated the six experimental classroom teachers on their implementation of the TEEM Program (Table II-6). A scale of 1-9 was used, the possible ratings ranging from "barely acceptable" through "average" to "outstanding". The highest ratings reflect the program coordinator's judgement that the teacher was successful in translating the TEEM approach into classroom activities and attitudes.

The ratings were used to rank classes on TEEM implementation. These ranks were then correlated with the ranks assigned the classes by assessment of the average gain achieved by children from pre- to posttesting on the McCarthy summary score (Table II-6). The latter data was not known to the program director who rated teachers. The rank-order correlation of .63 reveals a noteworthy relationship between successful TEEM implementation and children's cognitive gains as measured by the individually-administered, standardized test (Spearman rank-difference method, p=.63).

Solomon Four-Group Design

All children present in the six TEEM and two Comparison classrooms were tested in mid-April, 1972. This resulted in approximately twice as many children (99 TEEM and 29 Comparison) with valid posttest data on the McCarthy than with pretest.

The main effect of the pretesting experience on posttest scores was determined by using a Solomon Four-Group Design (Campbell and Stanley, 1967, pp. 24-25). Comparison and TEEM

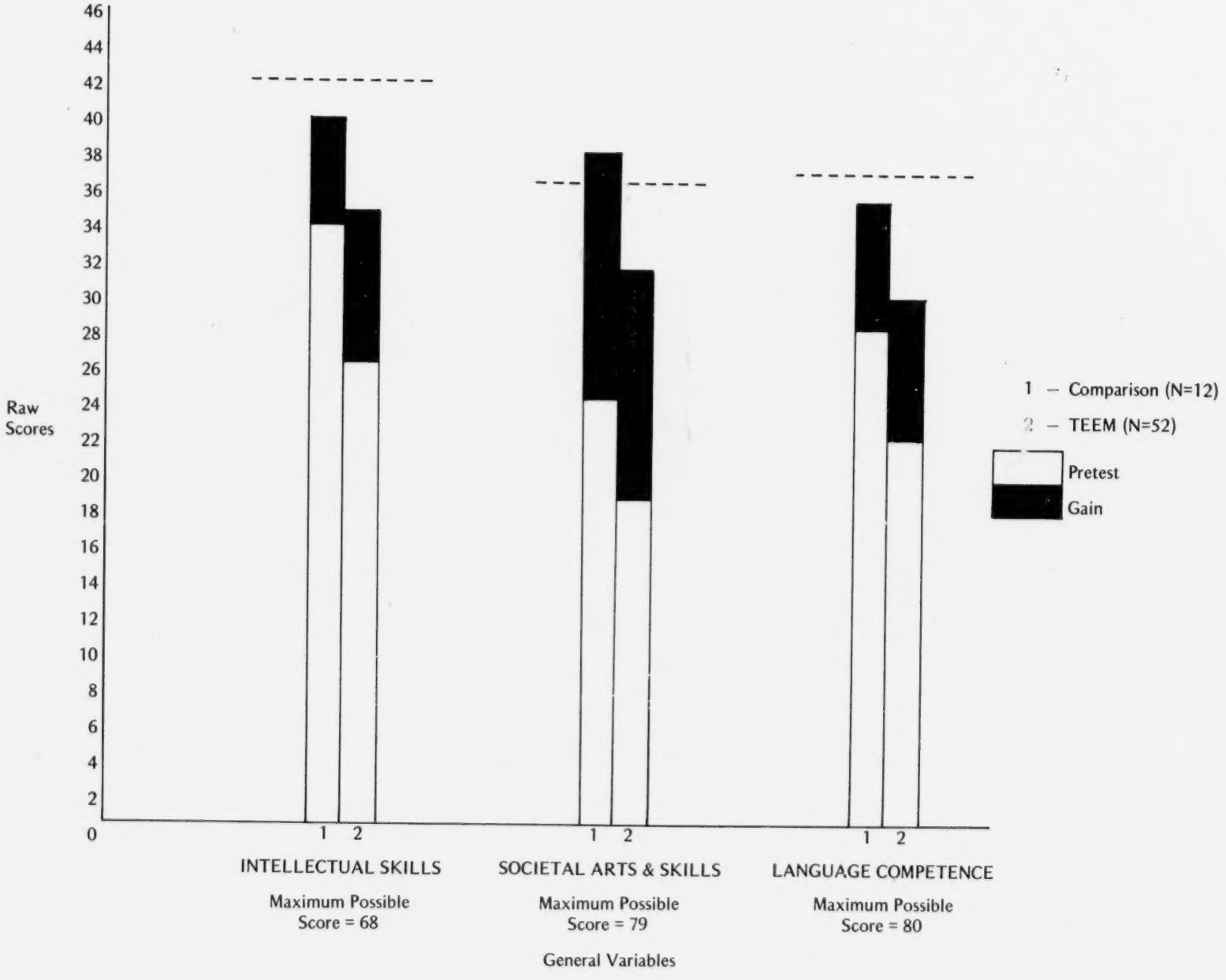


Fig. II-1. Pre- and posttest McCarthy raw scores on three general variables. Dotted line indicates normative scores for 5-year-olds.

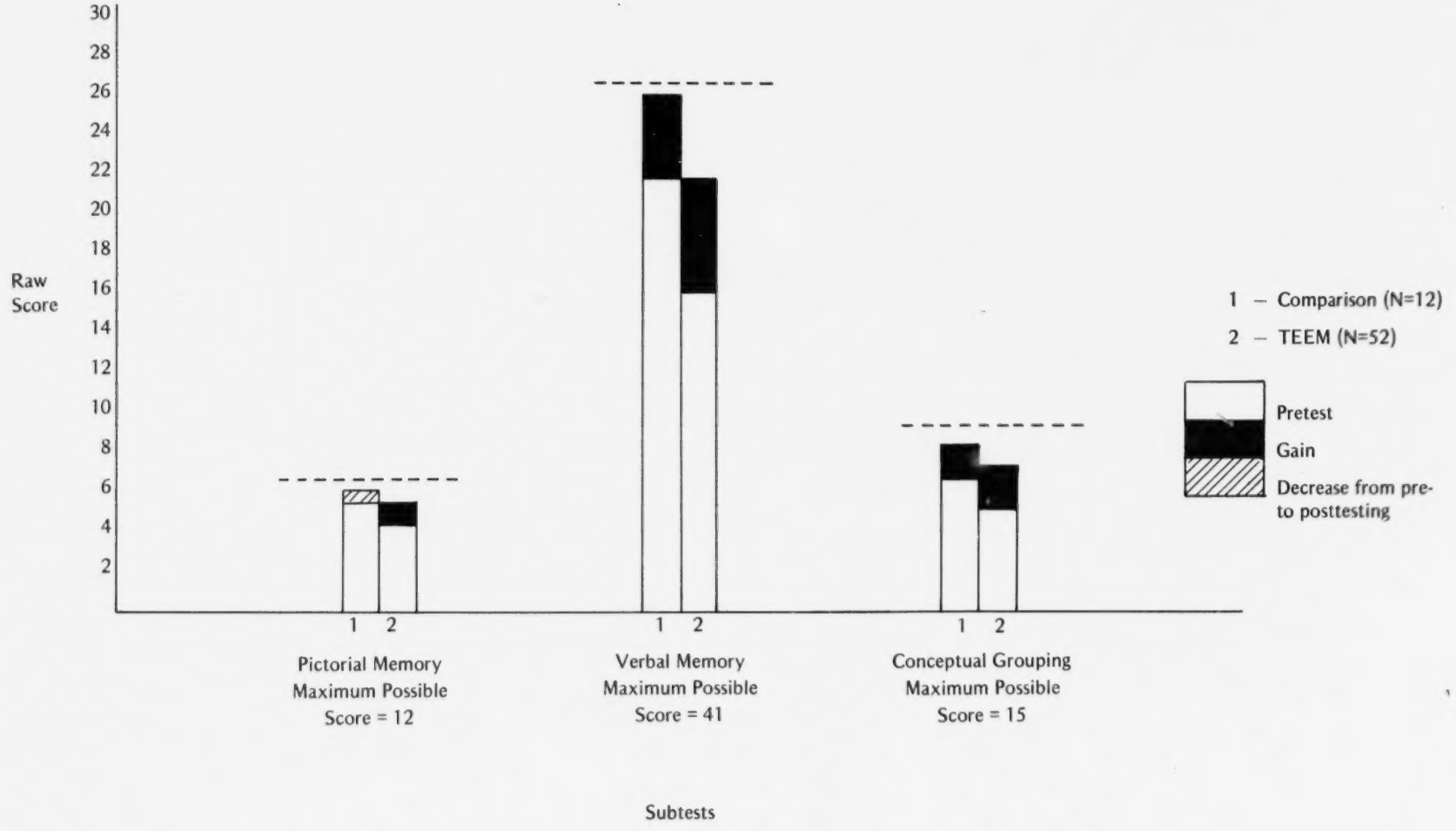


Fig. II-2. Pre- and posttest scores on McCarthy subtests included in the Intellectual Skills variable, Comparison vs. TEEM children. Dotted line indicates normative scores for 5-year-olds.

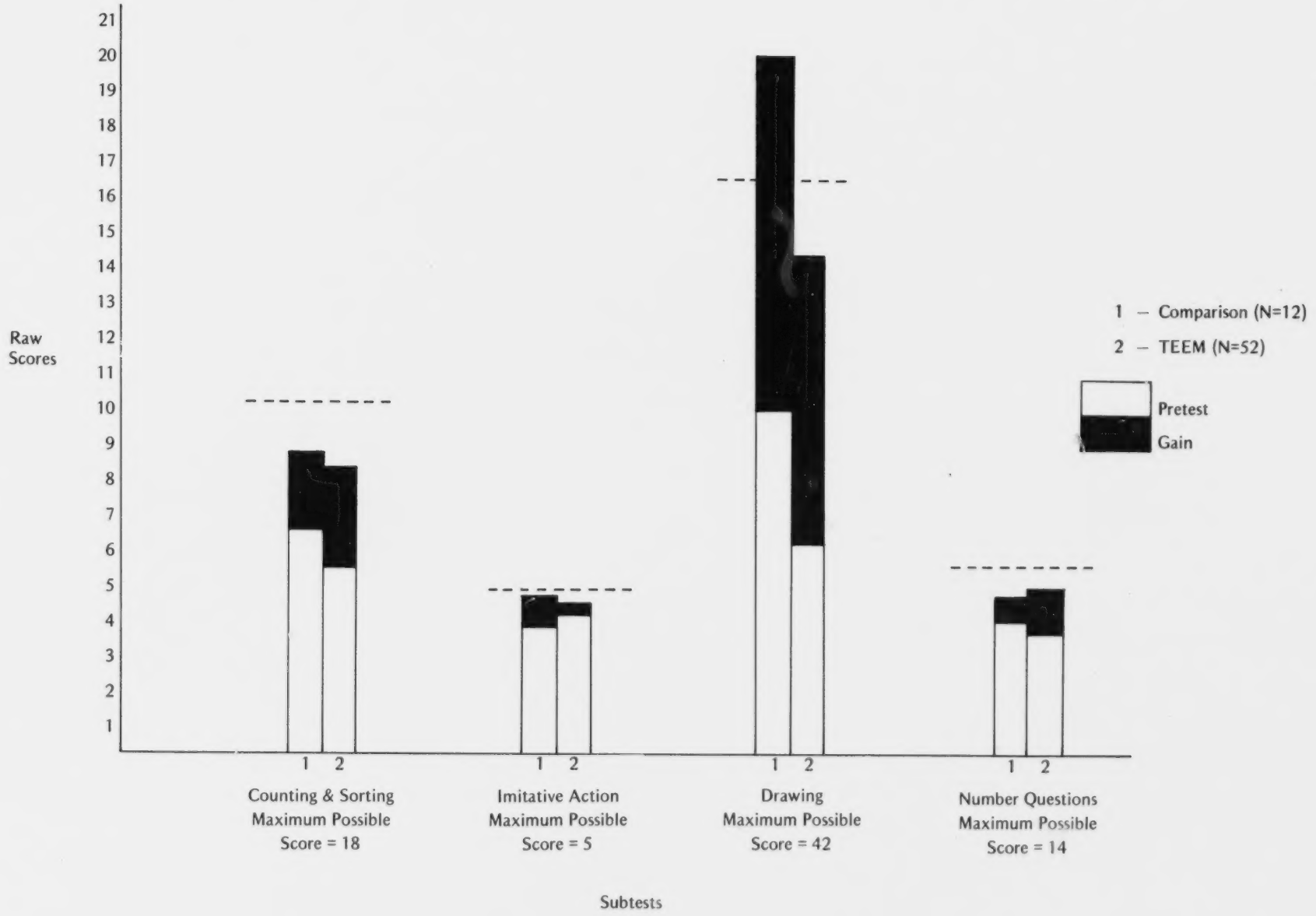


Fig. II-3. Pre- and posttest scores on McCarthy subtests included in the Societal Arts and Skills variable, Comparison vs. TEEM children. Dotted line indicates normative scores for 5-year-olds.

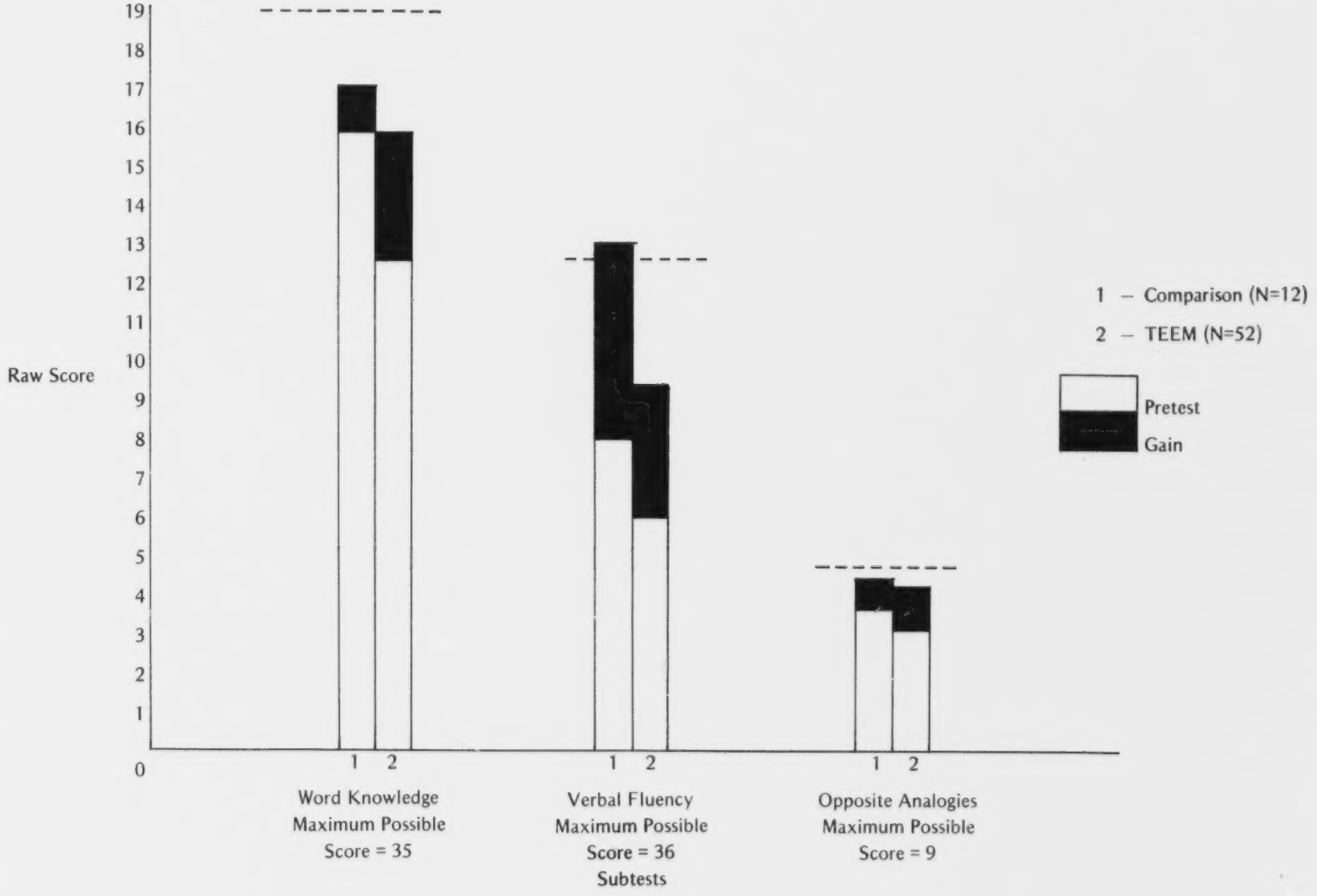


Fig. II-4. Pre- and posttest scores on McCarthy subtests included in the Language Competence variable, Comparison vs. TEEM children. Dotted line indicates normative scores for 5-year-olds.

TABLE II-6

RELATIONSHIP BETWEEN TEEM TEACHER RATINGS AND CHILDREN'S COGNITIVE GAINS

	TEEM Implementation: Teacher Rating Assigned by Program Coordinator	Classes Ranked by Teacher Effectiveness as Rated	Classes Ranked by Children's Mean Gain on McCarthy total, pre- to posttesting
School A	6	1.5	2
School B	6	1.5	3
School C	5	3	1
School D	3	4.5	5
School E	3	4.5	6
School F	2	6	4

Note.—Rank-difference correlation between last two columns: p=.63.

subjects who received pretests (groups one and two for this analysis) were not significantly different on either individual subtest or McCarthy summary posttest scores from Comparisons and TEEM children who received posttest only (group three and four). Nor was there an interaction between the pretesting factor and the control-experimental distinction. Therefore the effect of pretesting *per se* on the posttest performance was considered null.

Posttest Data

The addition of the second Comparison class at posttesting yielded a composite group (N=29) which was more comparable to posttested TEEM children (N=99) than was the case for the smaller pretested samples. For example, Appendix A reveals that demographic data on the total Comparison and experimental groups is less divergent than was true at pretest. More important, the doubling of sample size at posttest resulted in an identical age of children in the posttest Comparison and posttest TEEM groups. (Both groups average four years, 10.8 months at time of posttest).

Table II-7 indicates that TEEM children are slightly superior on seven of the 10 (70%) McCarthy subtests at posttest. These are Pictorial Memory, Conceptual Grouping, Counting and Sorting, Imitative Action, Number Questions, Word Knowledge and Opposite Analogies. On one of these (Number Questions) there is a statistically significant difference in favor of TEEM children at $p<.05$. The fact that Comparisons scored higher than TEEM children on all 10 of the pretest subtests, but TEEM surpassed Comparisons on seven of the 10 post-subtests, is certainly noteworthy despite the lack of statistically significant differences.

Means on the total posttest sample for the three TEEM goal areas of Intellectual Skills, Societal Arts and Skills, and Language Competence are virtually identical for the TEEM and Comparison groups, as Table II-7 indicates. Similarly, posttest McCarthy summary scores are very similar for TEEM children and Comparisons (means of 97.83 and 98.07, respectively).

Motivational Base

The instrument used in the assessment of the children's motivational base growth was the Schaeffer Behavior Inventory (SBI). (For a more detailed description of SBI, see page 7.)

As with any rating scale, the reliability and validity of this inventory of classroom behaviors is dependent largely on the child raters (in this case, teachers). In three of the six TEEM Head Start classes the teachers changed during the school year. Hence, nine different raters were involved in

TABLE II-7

McCARTHY POSTTEST MEANS AND F-RATIOS FOR TEEM AND COMPARISON HEAD START CHILDREN

	TEEM N=99	Comparison N=29	Test of Difference
	$\bar{X}$ Raw Score	$\bar{X}$ Raw Score	F-ratio
INTELLECTUAL SKILLS	41.23	41.38	1.
Pictorial Memory	5.96	5.14	3.46
Verbal Memory	21.51	22.66	<1
Conceptual Grouping	7.69	7.55	<1
SOCIETAL ARTS AND SKILLS	32.13	32.00	1
Counting and Sorting	8.12	8.03	<1
Imitative Action	4.28	4.21	<1
Drawing	14.85	15.49	<1
Number Questions	4.88	4.17	4.48*
LANGUAGE COMPETENCE	30.55	30.72	1
Word Knowledge	16.39	16.10	<1
Verbal Fluency	9.79	10.52	<1
Opposite Analogies	4.36	4.10	<1

*$p < .05$.

rating the TEEM children. Changes in the raters leave the data, especially the gain scores, in a suspect position. The most valid pre-post data therefore comes from the three TEEM classes where there was no teacher change. The same teachers completed the SBI protocols in the Fall and in the Spring. This is an important factor because individual differences in operational definitions of such subjective terms as "likes," "angry," "earnestly," "enjoys" will yield differences in ratings. Although individual differences among the three teachers who did not change from the pre- to post-ratings are expected, it would not be anticipated that their definitions of the behavior over the year would drastically change. The conjecture that these teachers may have utilized the inventory's points of emphasis as a working framework for classroom behavioral analyses during the school year is not unlikely.

The ratings are averaged by groups, some groups having more than one rater at different times. Differential ratings may result if the judgments were based on a six month, or a six week, sample of the children's behavior. Consequently, another factor to be considered is the length of time the new rater had been the classroom teacher.

Psychometricians and social psychologists acknowledge that this Likert-type of rating scale is notoriously subjective. One prevalent artifact is Experimenter (Rater) bias. Inventory results generally reveal the rater's own individual preferences and prejudices about what such a test should reveal. Another artifact of the Likert scale is response set or response bias. For example, a rater may fall into a middle-of-the-road set which prevails throughout his ratings almost regardless of the item content. The facts presented in the preceding three paragraphs should be considered as one analyzes the SBI data.

In view of these less than optimal methodological points, the reliability data is quite good. The reliabilities of the items on the SBI from pre- to posttesting are superior to those usually obtained

for individual items. This is even more noteworthy when consideration is given to the facts of a seven-month span between testing sessions and of mid-year teacher changes in three of the six classrooms. The range of the reliabilities of the 15 individual items for all six TEEM classes is between .40 to .59, with a mean reliability of .50. Considering just the three classes which experienced no teacher change for the school year, the item reliabilities range from .48 to .72, with a mean of .63. In the interest of clarity, the six combined TEEM classes will be referred to in this data analysis as group A and the three TEEM classes where the teachers did not change in mid-year as group B.

A summing of the items in the three major variables of Task Orientation, Extraversion and Hostility yields the subtotal reliabilities presented in Table II-8.

TABLE II-8

SCHAEFFER BEHAVIOR INVENTORY TEST-RETEST RELIABILITIES

	Group A: Children from all six TEEM classes N=82	Group B: Children from three TEEM classes with no teacher change N=36
Task Orientation	.63	.76
Extraversion	.53	.69
Hostility	.58	.67

Gain scores reveal numerical improvement from one testing session to another. The SBI gain scores for both group A and group B are presented in Appendix D, and are illustrated in Figures II-5 through II-8. Figure II-5 presented the pre- and posttest mean ratings received by the children on the five items included under the Task Orientation rubric. The teachers judged their children as being average ("half the time") or above ("frequently") on these items. Gain scores are statistically significant at $p<.05$ level for group A on two items: "Stays with a job until he finishes it," and "Becomes very absorbed in what he is doing." Consistently group B ratings are slightly higher. Decreases from the pre- to posttesting are seen for this group on two items: "Works earnestly at his classwork. Doesn't take it lightly," and "Watches carefully when a teacher or classmate is showing how to do something."

The ratings on five items clustered under the Extraversion variable are disclosed in Figure II-6. Overall, these judgments are somewhat higher than those of the previous figure. Group A experienced virtually no change from the pre- to posttesting on three items: "Likes to take part in activities with others," "Enjoys being with others," and "Seeks social contact with others." The change on one item: "Tries to be with another person or group of people," was statistically significant at $p<.05$ level, as was the rating change for group B on one item: "Likes to take part in activities with others."

Figure II-7 presents the mean teacher judgments on the Hostility items. As would be expected, the composite picture is one of lowered ratings, in the range of "never" to "occasionally," and negative gain scores. Both groups achieved the latter for one item: "Slow to forgive when offended." In addition, group B ratings decreased from pre- to posttesting on two other items: "Gets impatient or unpleasant if he can't get what he wants when he wants it," and "Angry when he has to wait his turn or share with others." The two rating sessions yielded virtually no change for group B on one item: "Complains or whines if he can't get his own way."

Overall analysis by item indicates three statistically significant gain scores ($p<.05$) for group A and only one for group B. On an analysis of fifteen items the latter finding is not superior to what would be expected merely by chance. On seven of the 15 items for group A and on nine of the 15 items for group B, the pre- to posttesting changes were in the appropriate direction (e.g., increases for Task Orientation and Extraversion items, and decreases for Hostility items).

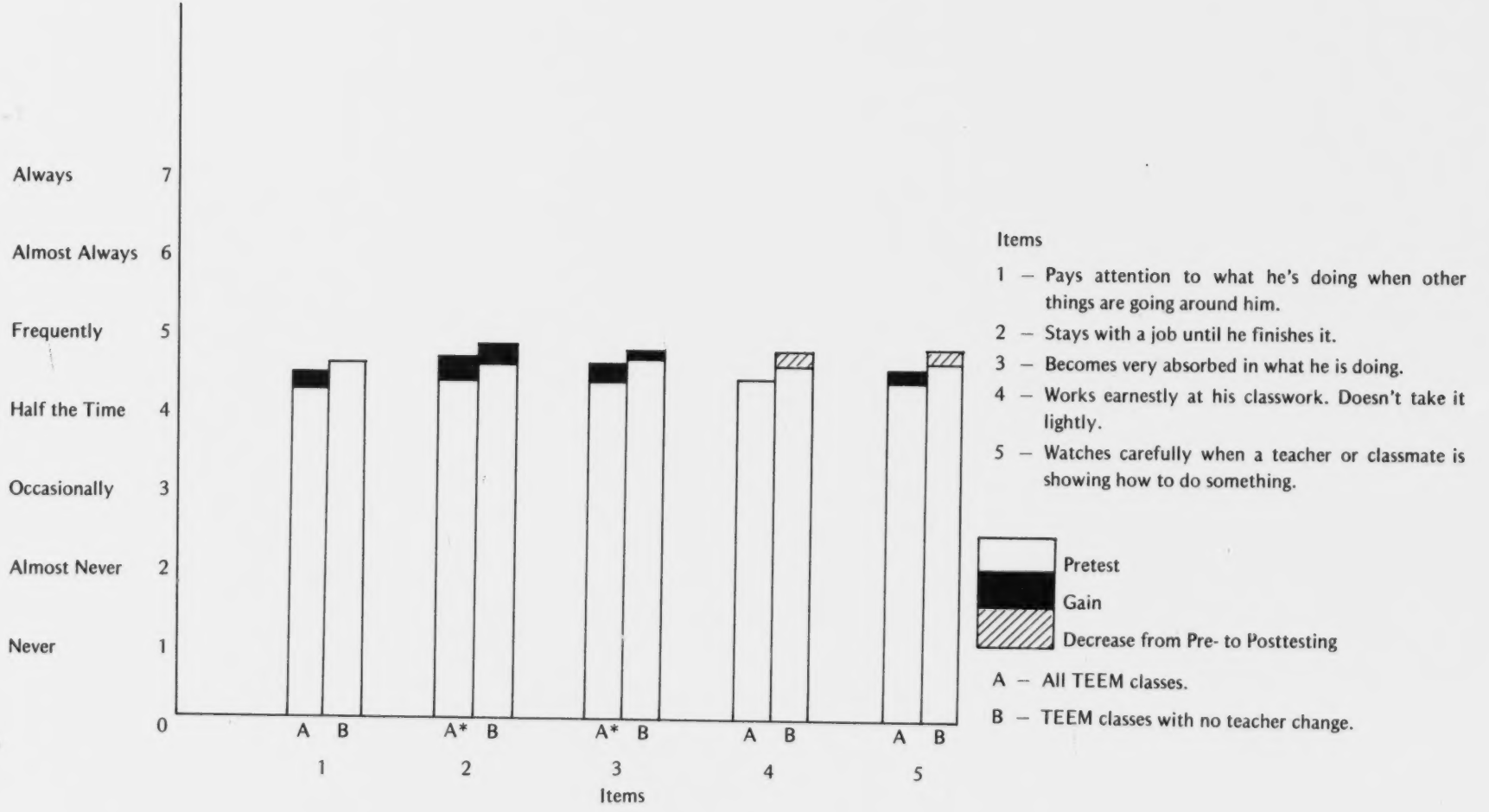


Fig. II-5. Pre- and posttest mean ratings on Task Orientation items for TEEM children as measured by the Schaeffer Behavior Inventory.

*p<.05.

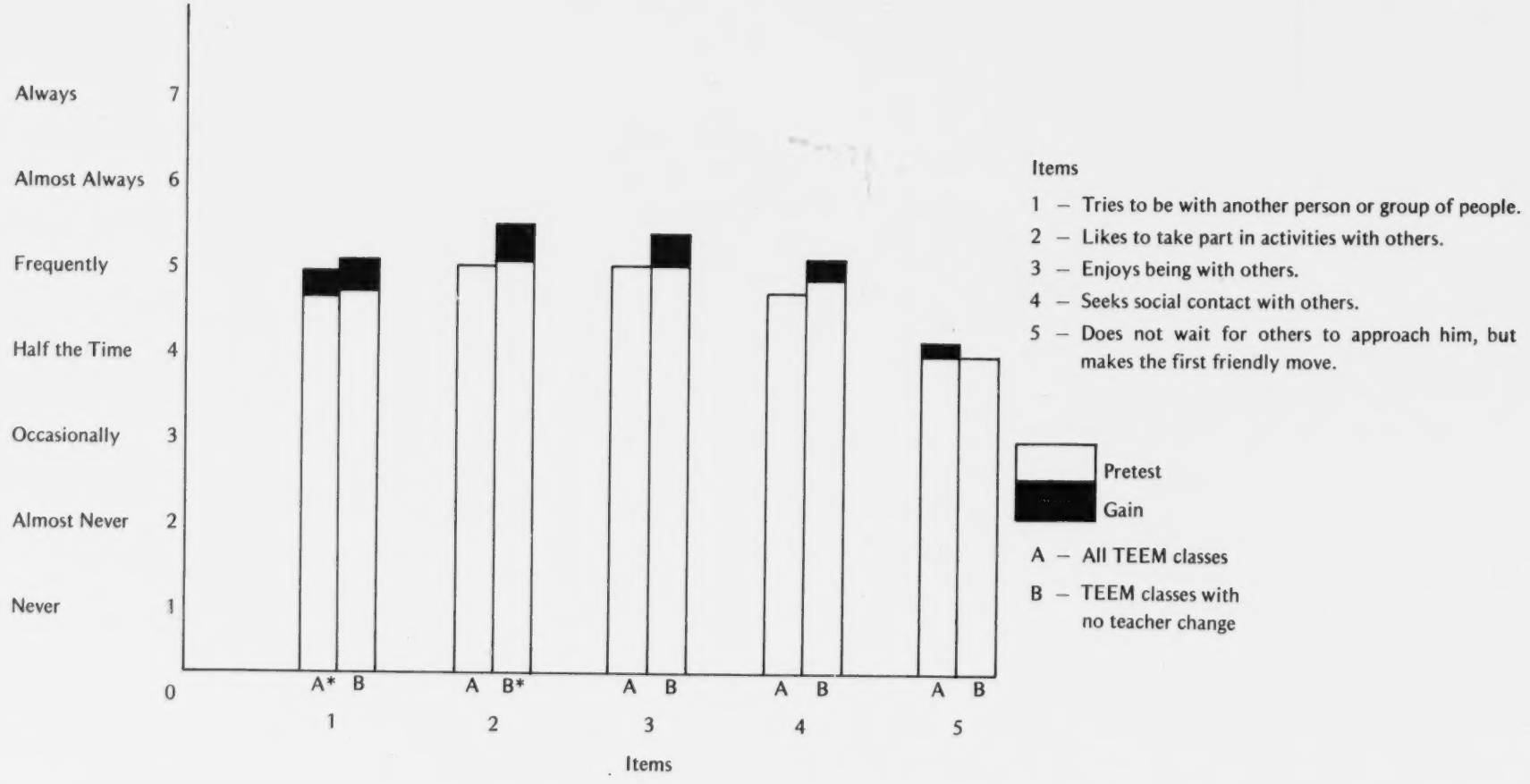


Fig. II-6. Pre- and posttest mean ratings on Extraversion items for TEEM children as measured by the Schaeffer Behavior Inventory.

*$p < .05$

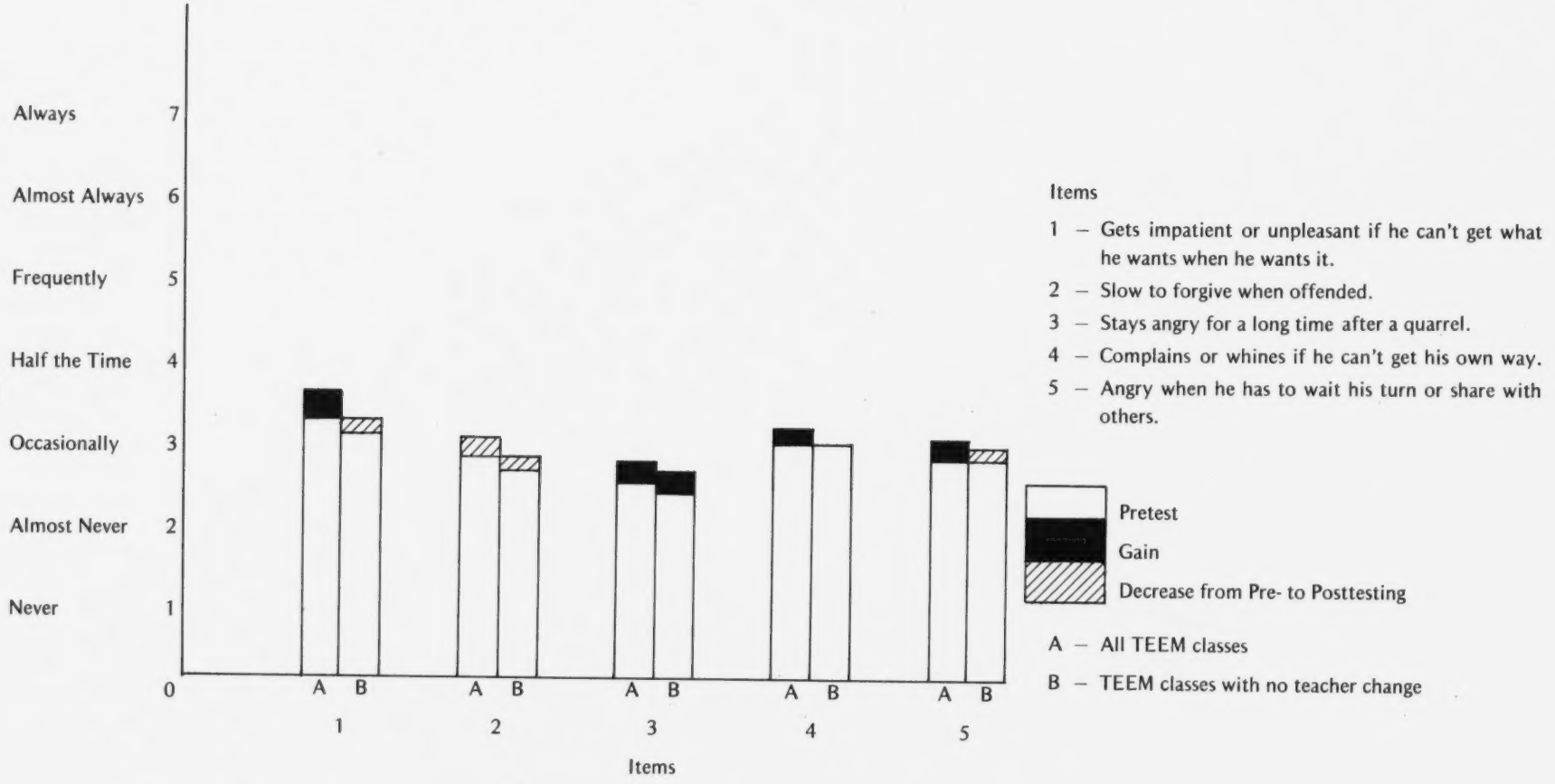


Fig. II-7. Pre- and posttest mean ratings on Hostility items for TEEM children as measured by the Schaeffer Behavior Inventory.

Trends on the three major variables of Task Orientation, Extraversion and Hostility are shown in Figure II-8. For both groups the means judgments are positive, with Task Orientation and Extraversion ratings ranging from "half the time" to "frequently," and Hostility ratings ranging from "occasionally" to "almost never." The direction of change is appropriate in two of these three categories for group A. Overall improvement from pre- to posttesting on all three clusters is seen in group A ratings. However, such an increase is desirable only for Task Orientation and Extraversion variables. Group B ratings did not appreciably change on Task Orientation and Hostility but did improve on Extraversion.

The two Comparison class teachers did not complete the SBI on their pupils in the fall. Therefore the only TEEM vs. Comparison group data available are on the posttest. The group means of these ratings for the TEEM and the Comparison children are presented in Appendix E. The mean ratings on the three basic behaviors are revealed in Table II-9.

TABLE II-9

SCHAEFFER BEHAVIOR INVENTORY POSTTEST MEANS FOR TEEM AND COMPARISON HEAD START CHILDREN

	TEEM N=104	Comparison N=26
Task Orientation	4.6	4.5
Extraversion	4.8	4.7
Hostility	3.1	3.4

There were no statistically significant differences between the groups on either the items or on the subtotals of the three major variables. The item ratings for the two groups were quite close, with a negligible rating difference. A comparison of the TEEM and Comparison group ratings presented in Table II-9 indicates a slight superiority in the desired direction in favor of the TEEM children on all three variables (Task Orientation, Extraversion and Hostility).

Classroom Process

Information on classroom process was gathered through observation techniques. Two different instruments were utilized in this phase. One was an experimentally developed technique designed to assess certain process goals of the TEEM program. The Classroom Attitude Observation System (CAOS) was a combination of observation instruments which included categories for *activity types*, incidence of *inappropriate behavior*, and *grouping* strategies. It includes an experimental modification in which the teacher is absent from the classroom with observation ongoing. The other instrument was a more traditional observation format developed by Stanford Research Institute as part of their national study of Head Start Planned Variation. The Classroom Observation Procedure (COP) has categories for *physical plant quality*, *activity types*, as well as *five-minute interaction sequences*. The analysis will incorporate a validity matrix in which information gathered by these two techniques will be compared.

The Classroom Attitude Observation Schedule data were reduced into summary variables for analysis. These variables focussed on key concerns of TEEM program implementation. The initial analysis looked at *mean group size*, which was calculated by dividing total number of children counted in a twelve-minute phase by the number of groups counted in that phase. A second variable was that of *mean number of children* engaged in an activity at any one time during the twelve-minute phase. In terms of the recording instrument, this was mean cell size and was calculated by dividing the total number of children counted in a phase by number of cells used in a phase. The third variable was *inappropriate behavior*. This was simply a total count of incidents of inappropriate behavior observed during the twelve-minute phase. Two other variables tested were *mean number of groups* in a phase, and *mean number of activities* for a given phase.

The above five variables were combined for TEEM classrooms and Comparison classrooms, and the group means over the 3 phases were analyzed with analysis of variance procedure. Results of these analyses appear in Table II-10.

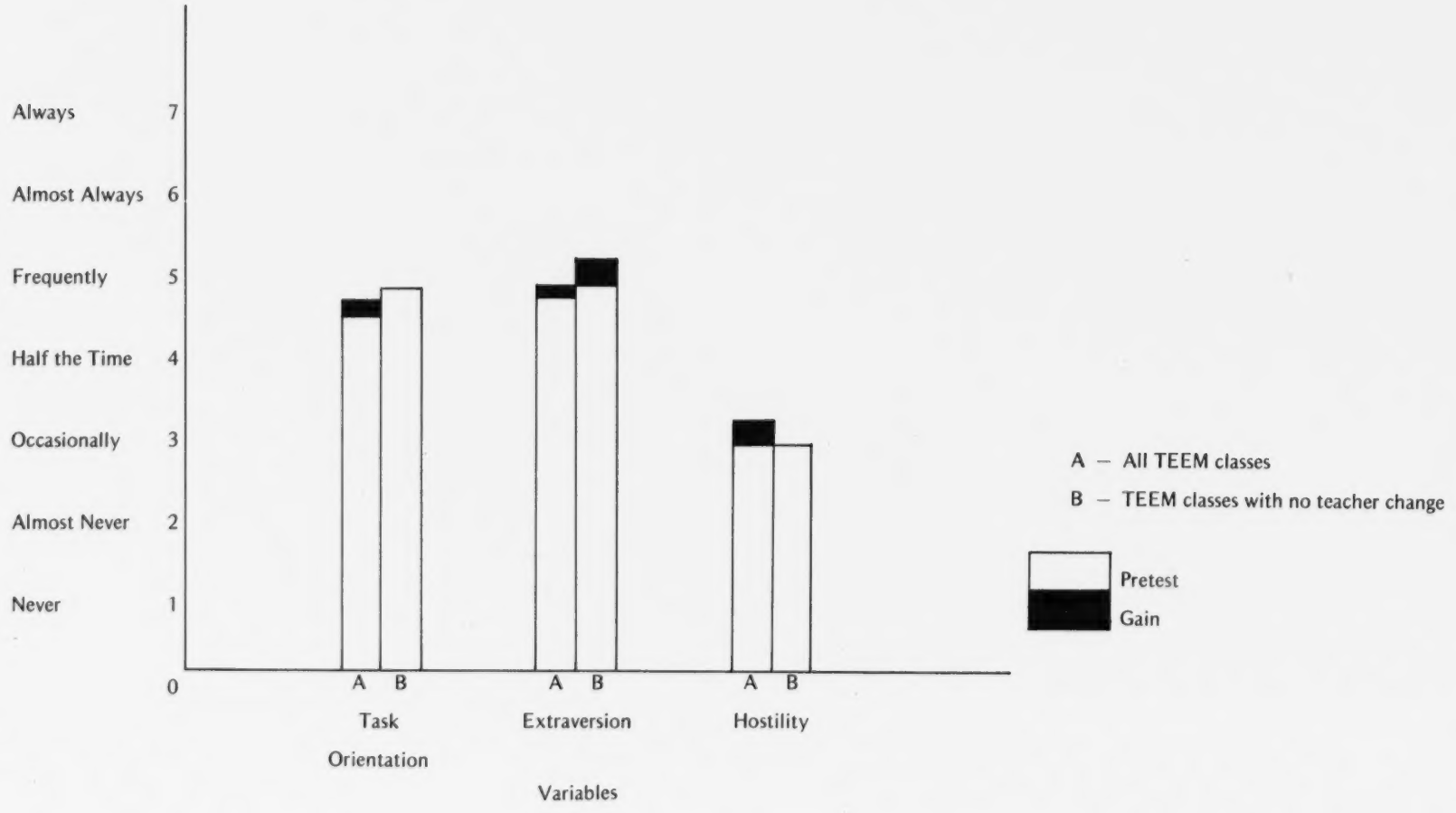


Fig. II-8. Pre- and posttest mean ratings on Task Orientation, Extraversion and Hostility as measured by the Schaeffer Behavior Inventory.

TABLE II-10

ANALYSES OF VARIANCE OF TEEM AND COMPARISON CLASSROOMS ACROSS PHASES FOR SELECTED CAOS OBSERVATION SUMMARY VARIABLES

Variable	Source of Variation	df	SS	MS	F
Mean Group Size	Group (TEEM vs. Comparison)	1	0.006	0.006	<1
	Phase	2	0.340	0.170	3.81+
	Group x Phase	2	0.076	0.038	<1
Inappropriate Behavior	Group (TEEM vs. Comparison)	1	193.39	193.39	20.80**
	Phase	2	279.75	139.87	14.36***
	Group x Phase	2	262.70	131.35	13.48*
Mean Cell Size	No significant results were obtained.				
Mean Number of Groups	No significant results were obtained.				
Mean Number of Cells	No significant results were obtained.				

+$p<.10$.

*$p<.05$.

**$p<.01$.

***$p<.001$.

Another analysis used independent ratings of teachers. The Lincoln Head Start Director was asked to rate all of the teachers on level of implementation (c.f. page 12). The lowest teacher rated was a "2" and the two highest were "6's". Since these ratings were assigned independently of the variables examined for CAOS, one high-rated teacher was selected from the two high rated teachers[5], and was compared with the low rated teacher on variables showing variation on CAOS. For further examination on the data, activity categories were collapsed into five broad types of activities. Type I consists of "traditional" academic activities and includes arithmetic, language, social studies and science. Type II consists of cognitive learning activities for young children but not in the "traditional" sense. Included in this type are stories, games, puzzles, arts, crafts, cooking, and building. Type III consists of play and role playing. Type IV consists of snacks and management. Example of management activities are: cleaning up, handing out material, going after supplies, sharpening pencils, teacher and child talking. These activities apply to both children and adults. Type V is non-focused activities, such as transitional activities, children out of the room, and wandering.

These five activity classes were used to examine the porportion of time spent by adults and children in each activity type during phases A and C (when both were in the room together). These data were examined for the same high-rated classroom and low-rated classroom (Figures II-9 and II-10).

Due to the exploratory nature of the research and the small sample of classrooms generous confidence limits were set for acceptance of statistically significant findings. The maximum level

[5]Of the two highest rated teachers, one taught the whole year in TEEM Head Start, while the second was a mid-year replacement. The former teacher was utilized in the analysis of the implementation outcomes.

acceptable was set at .10. It was felt that this level would give a sufficient indication that a variable should be explored in future research.

Of the five summary variables tested for analysis of variance, two demonstrated significance at acceptable levels (Table II-10).

Mean group size over all eight classrooms indicated differences across groups were significant ($p<.10$). There were no significant differences between TEEM implementated classes and Comparison classes.

The most conspicuous differences appeared with *inappropriate behavior.* Analysis of this variable indicated that all sources of variance were significant (groups: $p<.01$; phases: $p<.001$; group by phase: $p<.05$). Figure II-11 graphically describes this variable.

In comparing high and low rated classes, with only one classroom in each category, *inappropriate behavior* was not tested for statistical significance. However, a chart was prepared parallel to that for the summary variable, and the interaction pattern runs an essentially similar course (Figure II-12).

When adult participation was compared to child participation in the high and low classrooms, rank order correlation between adult presence and child presence in the various activity types was -.27 in the low rated classroom (Figure II-10). In the high rated classroom, the correlation between adult and child participation was .80 (Figure II-11).

The Classroom Observation Procedure (COP) was designed by SRI to: (1) assess the degree of implementation of classroom processes within various instructional models, and (2) assess child outcomes from the varying systems.

To achieve these assessment goals, the instrument is divided into three sections: Physical Environment Information, Classroom Checklist, and Five-Minute Observation. In the initial section, physical plant variables, such as lighting, noise level and seating arrangements are recorded. The second section entails a recording of what each adult and child in the classroom is doing, yielding a "snapshot" of the classroom. Several of these "data pictures" would then reveal the group size, adult focus, and activities typical in that class. In the third section the observer focuses on one individual, or group, and records in detail: the "who", "to whom", "what", and "how" of classroom interaction.

SRI-trained staff recorded a minimum of 24 observations per class in five Lincoln TEEM classrooms. Normative data on this instrument is not presently available. Consequently, to facilitate immediate utilization of the TEEM classroom observation data procured from SRI, the same observer was hired to make an equal number of observations in the two Lincoln Comparison classrooms.

Following the direction provided by the list of variables developed by SRI, the frequencies of 35 separate variables were compiled from the classroom observation data. For data analysis and graphic presentation purposes, these variables were grouped into the following variable constructs:

adult instruction
child-initiated learning
adult attention-five children or less
child self-expression
child-child interactions
positive behavior
negative behavior

A complete list of the COP variables included in these variable constructs are presented in Appendix F. Generally, both child and teacher behavior is included in a variable construct unless the variable construct title indicates otherwise.

In the interest of brevity, the data from only three of the seven classrooms are presented. Of these two are TEEM classrooms, those rated the highest and lowest by the Head Start Director of the teachers' success in implementation of TEEM (see page 12), and one is a Comparison classroom. In the majority of the variable constructs, the trend for the two TEEM classrooms is the same, and is different from the Comparison classroom. Appendix G provides the numerical frequencies for the three classrooms on the eight variable constructs.

Figure II-13 depicts histograms on the two variable constructs included in the Cognitive category. Both TEEM classes have a lower frequency over 24 observations of adult instruction, and a higher frequency of child-initiated learning than the Comparison class.

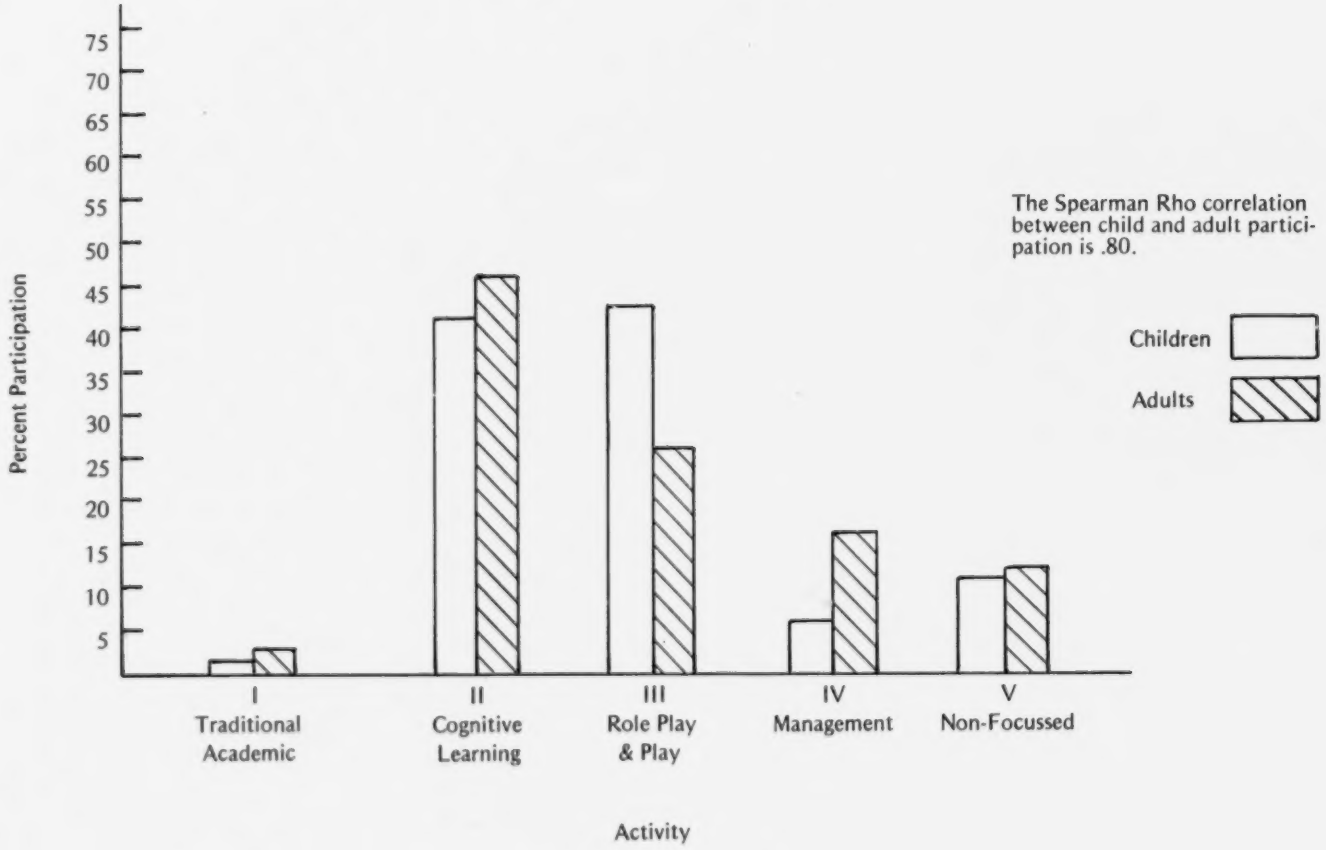


Fig. II-9. Location of adults and children by activities in high-rated classroom.

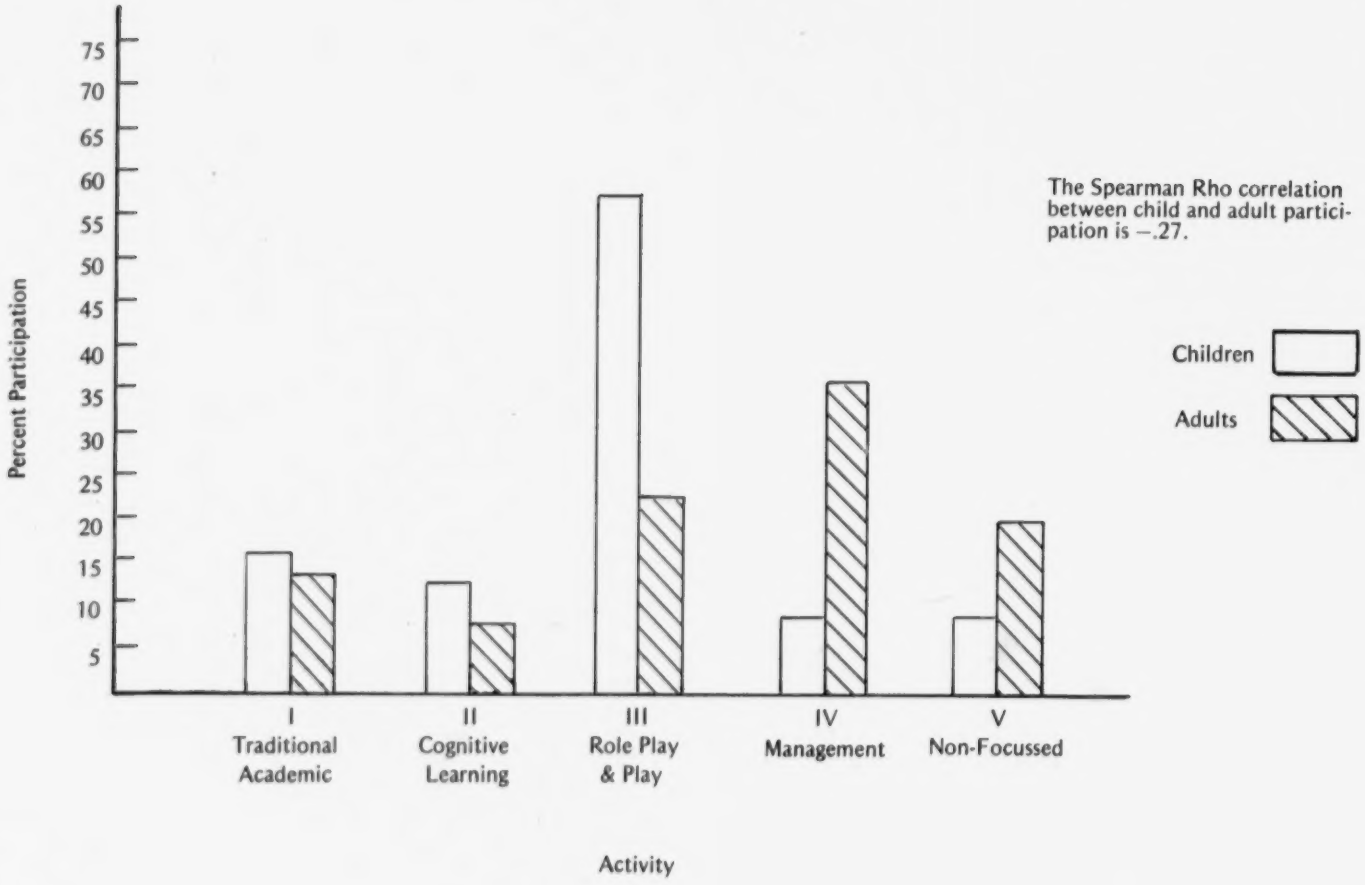


Fig. II-10. Location of adults and children by activities in low-rated classroom.

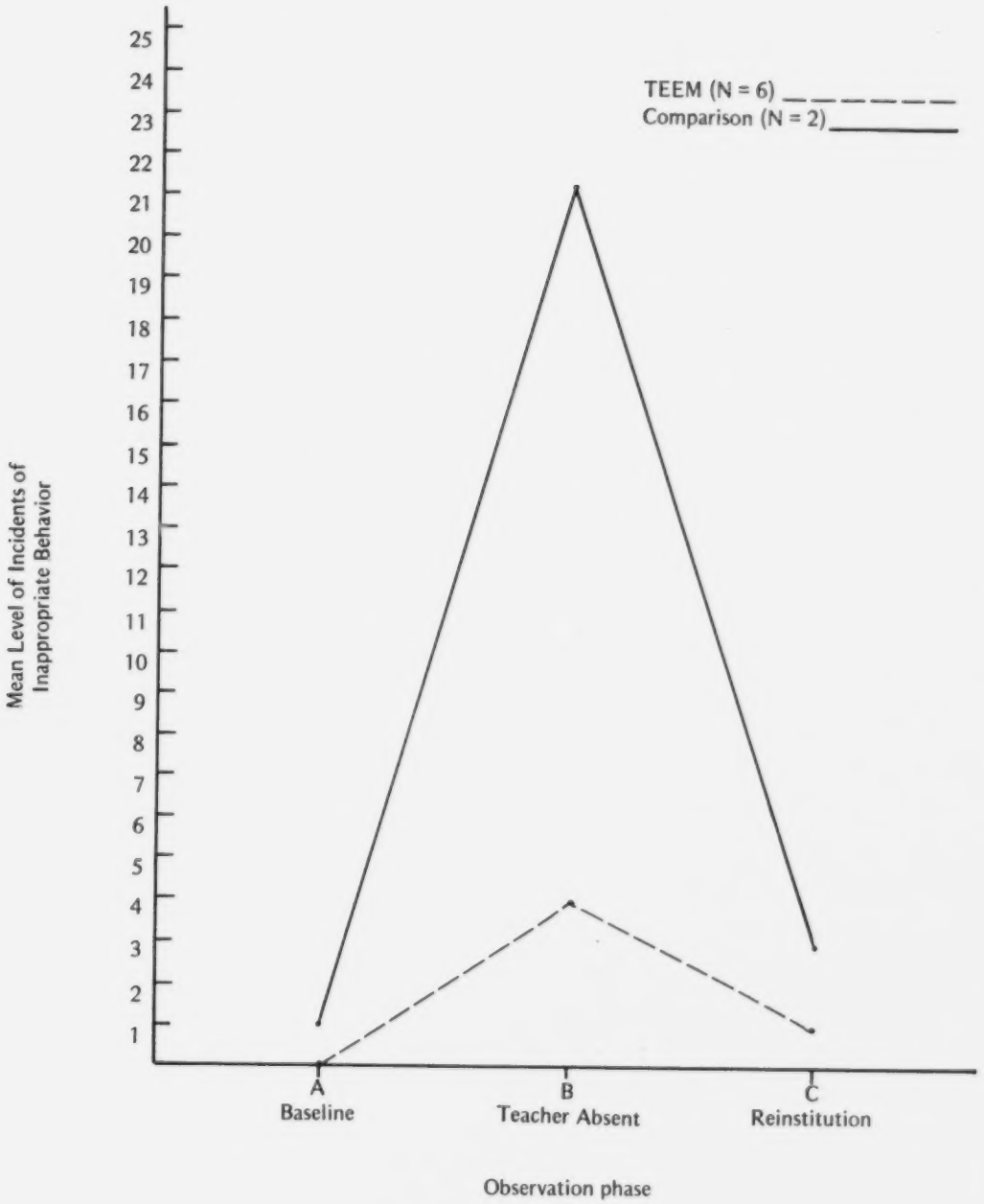


Fig. II-11. Incidents of inappropriate learning behavior in TEEM and Comparison classrooms across observation phases.

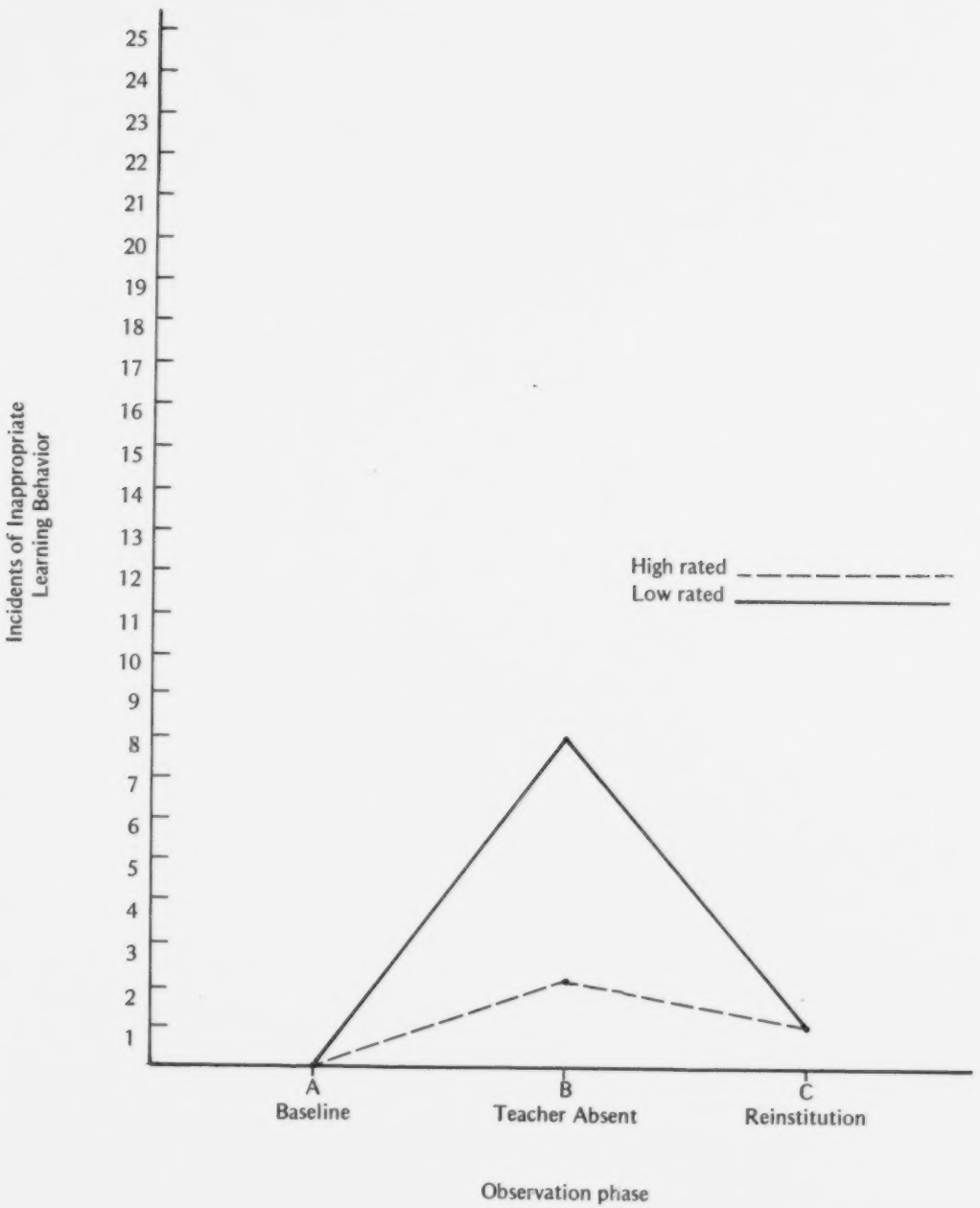


Fig. II-12. Incidents of inappropriate learning behavior in high- and low-rated classrooms across observation phases.

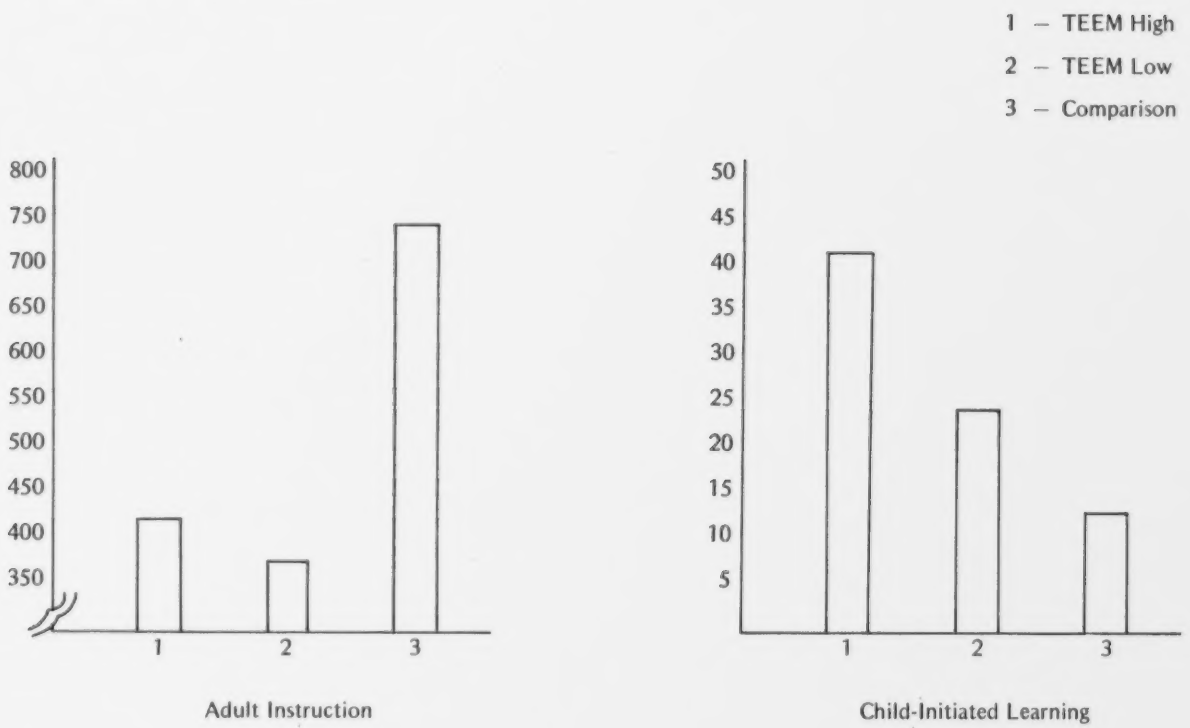


Fig. II-13. Individual histograms of the Cognitive variable constructs from the Classroom Observation Procedure data.

Figure II-14 presents individual histograms for the six variable constructs included in the Socio-Affective category. Overall, the trends indicated that in the TEEM classroom there was a lower frequency of adult attention–five children or less, more child self-expression, more single child-adult and more child-child interactions, more positive, and less negative behavior. The results of chi square analyses between TEEM High and the Comparison group, and TEEM Low and the Comparison group are listed in Table II-11. In the former analysis, 4 of the 8 comparisons are significant ($p<.01$). In the second analysis, 6 of the 8 of the comparisons are significant (2 of these $p<.05$; 4 of these $p<.01$).

Figures II-15 and II-16 illustrate a profile of the same variable constructs in the Cognitive and in the Socio-Affective categories for the three classrooms. The wide range of frequencies (see Appendix G) necessitated the use of two- and three-cycle semi-logarithmic presentation. When examining these two summary figures, it is well to keep in mind that this type of display minimizes differences in higher frequencies and maximizes those in the lower frequencies. Also, connecting lines between points representing the variable construct frequencies are drawn only to present a summary profile and are not intended to indicate a continuous variable on the abcissa.

Even with the present absence of normative data, these data presented reflect the usefulness of this instrument in determining objectively the ongoing processes within a classroom.

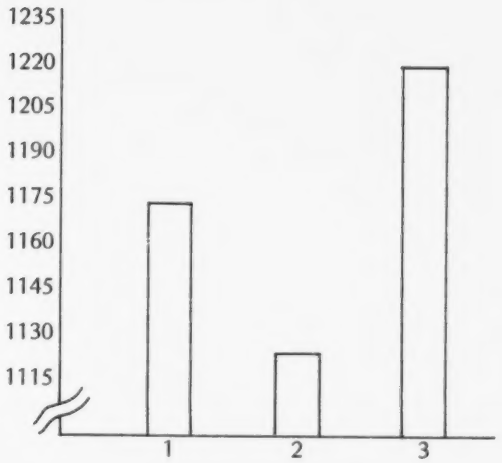
TABLE II-11

CHI SQUARE ANALYSES OF CLASSROOM OBSERVATION PROCEDURE VARIABLE CONSTRUCTS

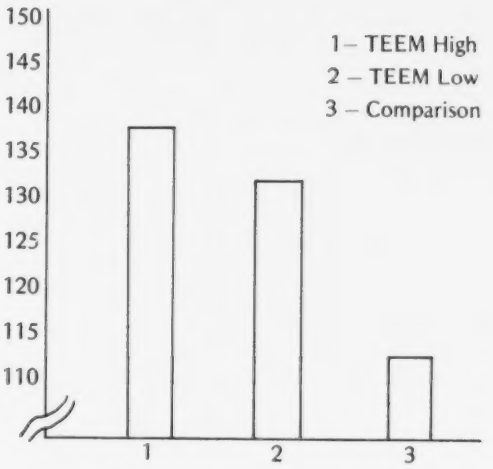
Variable Constructs:	Chi Squares:	
	TEEM High vs. Comparison Class	TEEM Low vs. Comparison Class
Adult Instruction	98.685**	134.759**
Child-initiated Learning	14.978**	3.613
Adult Attention: five children or less	.967	4.182*
Child Self-expression	3.034	1.889
Single Child-Adult Interaction	84.450**	32.246**
Child-Child Interaction	17.866**	9.413**
Positive Behavior	2.906	6.265*
Negative Behavior	.033	13.131**

**$p<.01$.

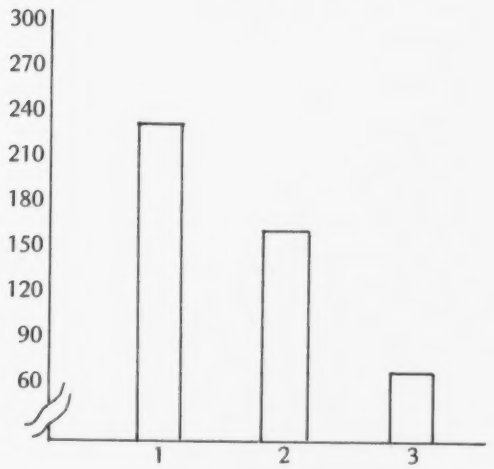
*$p<.05$.



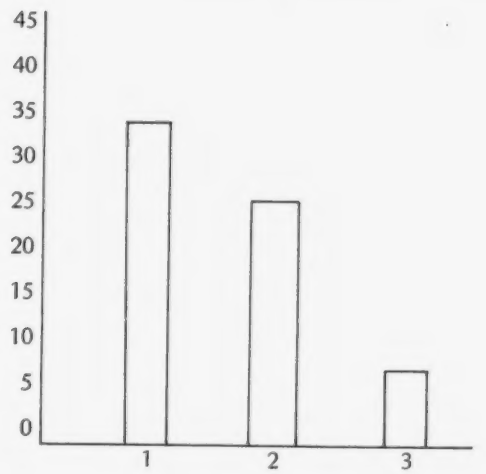
Adult Attention five children or less



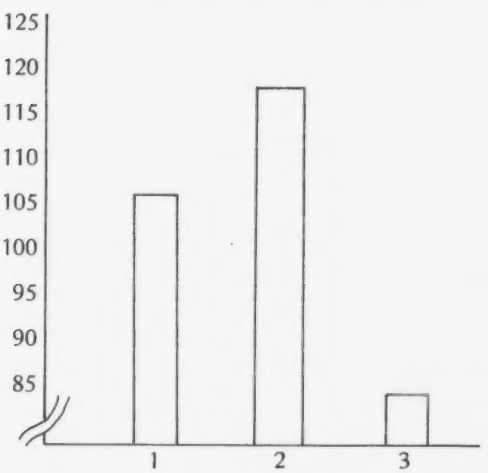
Child Self-Expression



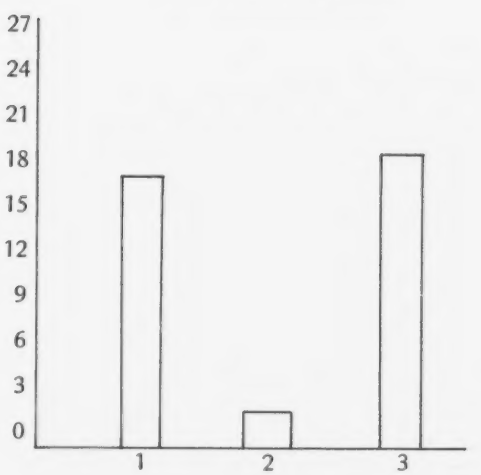
Single Child-Adult Interactions



Child-Child Interactions



Positive Behavior



Negative Behavior

Fig. II-14. Individual histograms of the Socio-Affective variable constructs from the Classroom Observation Procedure data.

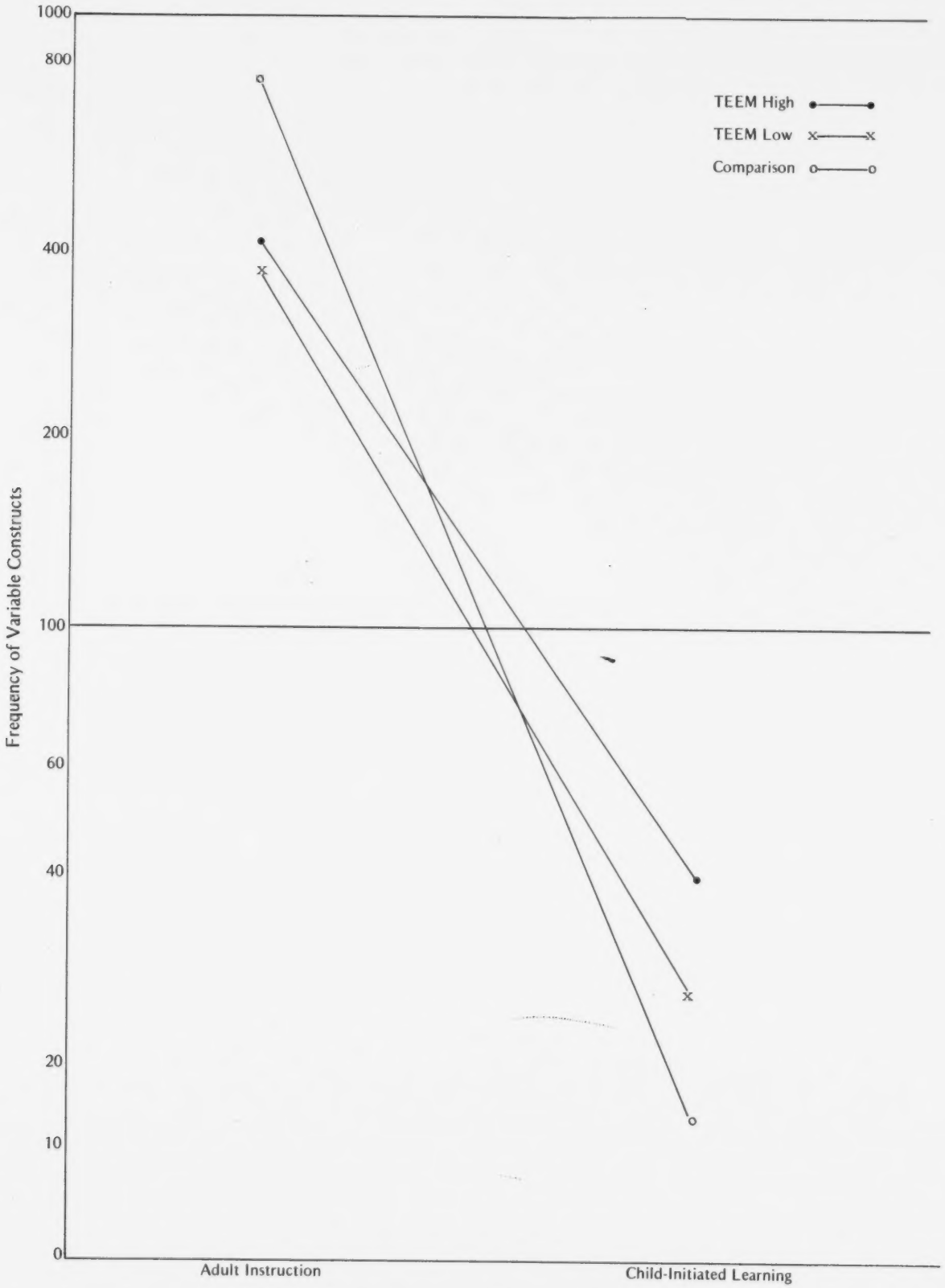


Fig. II-15. Summary profile of Classroom Observation Procedure variable constructs included in the Cognitive category.

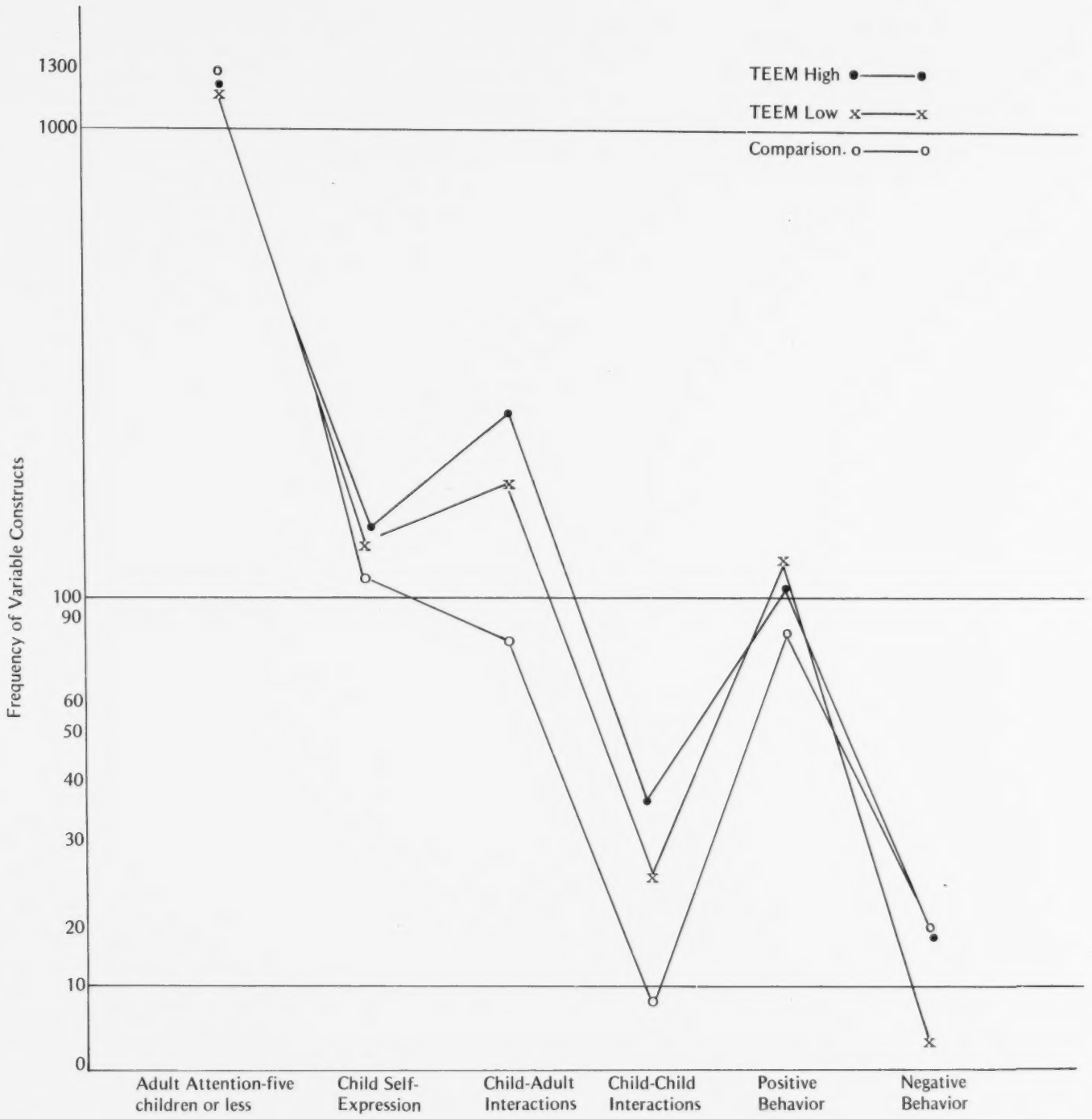


Fig. II-16. Summary profile of Classroom Observation Procedure variable constructs included in the Socio-Affective category.

DISCUSSION

The following discussion will focus on the stated goals of this study. These goals were concerned with the development of sponsor-specific outcome evaluation measures. In addition, the data were submitted to non-directional statistical analyses to test for differences between the group of children in the Planned Variation TEEM Head Start classrooms, and a Comparison group in locally implemented Head Start classrooms.

One goal of this study was the elaboration of an observational-evaluation system for the assessment of process goals within the TEEM classroom. This goal was met with the development of the Classroom Attitude Observation Schedule (CAOS). The CAOS technique represented a meld of methods taken from the experimental laboratory studies of children with current methods of observing social interaction in the classroom. The modification of the classroom environment by removing all adults (other than the observer) provided a crucial test of the implicit control system within the Head Start classroom.

The results from the CAOS observation sequence indicated quite striking differences between the TEEM classroom style and the locally implemented style. These differences were manifested despite the fact that the teachers in the locally implemented classrooms participated, at their discretion, in the TEEM training program offered through the Planned Variation funds. However, these Comparison classrooms did not receive the supplementary equipment and activity funds provided through Planned Variation participation. The conspicuous finding was the relationship between CAOS "teacher absent" phase and the manifestation of "inappropriate behavior" sequences in young children. The data indicate that, in the TEEM classroom, children are used to self management, and continued in their learning experiences with the teacher absent. In contrast, children in the Comparison classes manifested large increases in this dimension of non-adaptive behavior during teacher absence.

What explanation can be offered for these findings? The pioneer work of White and Lippitt (1960) has been influential in setting the direction of the "open education" movement in this country. Their research strategy was very similar to the one employed with CAOS, in which natural groups were unobtrusively observed, and information collected on leadership style and group development. Their studies looked at the relationship between leadership styles in informal boys clubs, and behavior patterns in participants. They found a strong relationship between democratic leadership style and self-directed behavior in the boys. In contrast, the autocratic leadership style produced boys who were more authoritarian and hostile.

It would not be untenable to assume that these kind of response styles can have their origin in the early school experiences of children. The data collected with CAOS provide information on the impact of the open system of early education in effecting self-directed, independent learning styles in young children.

Another set of related CAOS findings concerned the activity cycles of children and teachers in two TEEM classrooms, one rated as "high-implemented" and other rated as "low-implemented". Early experience in the task of widespread field delivery of the TEEM indicated that a wide range of classroom systems evolved, some high quality, most average, and a few unacceptable. In fact, the early findings of the SRI national evaluation study of Head Start indicated that there was often as much "program variance" within model sponsor sites as between different sponsor sites. With this in mind, a closer look was taken at the range of outcomes with the six TEEM classrooms. The "low-implemented" classroom showed adults predominantly concerned with management functions, while the children were involved in play sequences. In contrast, the "high-implemented" classroom showed adults involved in child-learning activities, and the children were engaged in cognitive-learning sequences. With a limited range of six classrooms, the strength of these differences must not be over-emphasized. However, the information does indicate that a rigorously applied evaluation system can find extensive amounts of information unavailable to a casual observer.

Another pervasive question in the evaluation of early education systems has been the relationship between level of model implementation and cognitive outcomes in children. The former construct refers to the fit between an idealized classroom (in this case, a TEEM classroom) and the range of classroom styles seen in the real-world field sites. One of the first priorities in developing a better understanding of the TEEM system in Head Start was the specification of the characteristics of the classroom system. This task was undertaken by members of the field instructional staff during 1971 at the request of the national evaluation contractor. These guidelines were utilized by the Director in Lincoln in her rating of the level of

TEEM implementation. On a nine-point Likert scale, the range of rated classrooms was two through six. No classrooms were described as near ideal criterion, while one was rated near the bottom.

The specification of a program-valid set of child measures was another goal of this study. Relating the classroom mean gains on the McCarthy summary score to the implementation rating produced a rank-over correlation of .63. With this small sample, this is not a significant relationship, but does indicate the direction for a future replication study. A conservative interpretation of this finding is that the conditions that facilitate cognitive growth in young children share some commonality with the expressed goals of the TEEM implementation system.

Additional information on the relationship between classroom process and child outcomes was offered through utilization of the data base collected by Stanford Research Institute (SRI). Economic and time limitations prevented complete utilization of this data source. However, information gathered on SRI's Classroom Observation Procedure was processed in time to be included in this report. This instrument shares some common information with the CAOS system, as well as unique data on the "actors" in the classroom and the nature of their interaction. A sample of only six classrooms does not permit direct statistical comparison of these two sets of information. However, a visual scan of the tables presenting these two data sets indicates some commonalities.

One of these relationships indicates that the occurrence of Single Child-Adult interaction was highest in the TEEM-high rated class, followed by TEEM-low rated and Comparison. This finding relates to the CAOS data which showed teacher and children learning together in the TEEM-high rated classroom. As well, the incidence of Child-Child interaction was highest in TEEM-high rated followed by TEEM-low rated and Comparison.

These similar findings, collected on two different observation instruments by different research groups, provide concurrent validity information on both the instruments and on the operation of the TEEM classroom. These data are currently being submitted to further analyses with the "teacher-child interaction" data base to explore these relationships.

Another goal of this study was the development of a set of child outcome measures that were congruent with the overall goal structure of the TEEM program. A major problem experienced in the SRI national evaluation study of Planned Variation, has been the specification of a test battery that is economical, valid and representative of the wide range of sponsor goals. Some sponsors rest content with the Stanford-Binet Intelligence Test or the Wide Range Achievement Test as an acceptable outcome measure. Because of the complex goal structure of the TEEM, no single available test was considered appropriate as an outcome measure. This led to the strategy of selecting different measurement tools for assessing the multiple outcomes of TEEM. The major source for the child data gathered in this study consisted of selected subtests from the recently published McCarthy Scale of Children's Abilities. In addition, teacher ratings were collected to assess children's growth in the motivational base area.

In terms of their attractiveness to children, this selected battery proved very successful. Children were stimulated by the variety of items, and none of the tasks presented serious entry problems to these Head Start children. The measurement properties of the tests were adequate, except for one subtest in which many subjects topped out (Imitation). The results indicated differential gain patterns between the TEEM and Comparison groups. In two of ten instances, these gains were statistically significant, and indicated greater growth for TEEM children. This finding, combined with the previously mentioned relationship between TEEM implementation and child gains, gives evidence for the validity of this battery for assessing TEEM outcomes.

Summary

The most important question facing the decision-makers at the Office of Child Development concerns the efficacy of the additional funds provided through the Planned Variation study. In terms of delivery of the TEEM program in PV Head Start, these funds were focussed in the teacher training effort, as well as provision for additional equipment and supplies in the classroom environment. The explicit question then relates to whether these additional resources were justified in terms of demonstrated differences in the children, teachers and community who receive Head Start services.

This study was explicitly directed to child outcomes, and the information presented here indicated that a quite different system of classroom organization existed in these two sets of classrooms. In the TEEM classrooms, children were oriented toward more independent learning sequences, and teachers participated in these experiences on an individualized basis. In contrast, the Comparison classes were typified by children in large group learning sequences, and there was evidence that more severe controls were exerted in classroom management. Similar information was produced through two independent sets of classroom observation data. If the values of participatory democracy, and the attendant self-directed learning styles, are important goals for young children, then some indication is given here that participation in a TEEM Head Start classroom contributes to that end.

Another set of data was concerned with cognitive growth as measured by a selected battery of psychological tests. The data gave some indication that TEEM children manifested an accelerated growth curve (in terms of McCarthy summary gain scores) on these tests, as compared to the Comparison group. However, only in a few instances were these differences statistically significant. If these goals are valued within the overall goal structure of the national Head Start program, information herein indicates TEEM has the potential to influence these outcomes.

The information indicated a wide range of classroom style variance evolves through the TEEM delivery system. Further, the range of outcomes was related to the quality of that implementation within the classroom. However, the short time frame of this study did not permit further examination of methods to bring all classrooms into an acceptable range of implementation quality.

Lacking direct information on the development of teacher competence within the TEEM delivery system, or the delivery of the parent involvement system, no comments can be made on the relationship of these aspects of TEEM delivery to child outcomes.

Future Directions

The study reported here, with limited resources, provided a minimum of solid findings and a raft of information to use in charting new research directions for TEEM. One direction is self-evident; classroom observation procedures provide the necessary tool for exploration of the important, but heretofore methodologically elusive, process goals of the open classroom. New techniques, such as CAOS, permit investigation of socio-emotional outcomes that have been impossible to pursue with traditional paper-and-pencil techniques.

Another avenue of research potential is to follow this group of children through their school experiences within the Follow Through feeder system. This longitudinal design has been developed, but additional funds have not yet been allotted.

This study has touched the surface of a number of important issues in the design of evaluation systems for compensatory programs. The great mass of unanswered questions still remains as a continuing challenge to the field.

APPENDICES

APPENDIX A

DEMOGRAPHIC DATA ANALYSIS

Pretest Sample Means and F-Ratios for TEEM (N=52) and Comparison (N=12) Head Start Children					
Demographic Variable	Valid N		TEEM	Comparison	Difference
	TEEM	Comparison	$\bar{X}$	$\bar{X}$	F-ratio
Family Income in $100's	51	10	38.31	43.50	<1
Employment Status of Head of Household[1]	52	12	2.12	1.83	<1
Years of Education of Head of Household	51	3	11.09	12.33	<1

Posttest Sample Means and F-Ratios for TEEM (N=99) and Comparison (N=29) Head Start Children					
Demographic Variable	Valid N		TEEM	Comparison	Difference
	TEEM	Comparison	$\bar{X}$	$\bar{X}$	F-ratio
Family Income in $100's	80	22	38.39	41.50	<1
Employment status of Head of Household[1]	81	25	2.17	1.92	<1
Years of Education of Head of Household	78	9	11.04	11.11	<1

[1] Encode system follows: 1 = Employed full time; 2 = parttime; 3 = seasonal; 4 = unemployed.

APPENDIX B

ANALYSIS OF COVARIANCE SUMMARY

Analysis of Variance of Pretest (x) and Posttest (y) McCarthy Summary Scores, Taken Separately							
Source of Variation	df	SS_x	SS_y	MS_x	MS_y	F_x	F_y
Among Groups (TEEM vs. Comparison)	1	4560.85	2994.69	4560.85	2994.69	8.24**	4.45*
Within Groups	62	34304.59	41703.74	533.30	672.64		

Analysis of Covariance on Posttests (y) Adjusted for Pretest (x) McCarthy Summary Score Differences							
Source of Variation	df	SS_x	SS_y	SS_{xy}	$SS_{y \cdot x}$	$MS_{y \cdot x}$	$F_{y \cdot x}$
Among Groups (TEEM vs. Comparison)	1	4560.85	2994.69	3695.72	82.56	82.56	<1
Within Groups	61	34304.59	41703.74	32710.22	10513.80	172.36	

Table of Adjusted Posttest Means				
Groups	N	Prestest $\bar{X}$	Posttest $\bar{X}$	Adjusted Posttest $\bar{X}_{x \cdot y}$
TEEM	52	67.84	98.56	102.43
Comparison	12	89.17	116.08	99.33

*$p < .05$.

**$p < .01$.

APPENDIX C

NORMATIVE McCARTHY SUBTEST MEANS

	Age in Years		
	4	4½	5
INTELLECTUAL SKILLS			
Pictorial Memory	5.4	5.8	6.6
Verbal Memory	19.9	23.6	27.9
Conceptual Grouping	5.9	7.1	9.4
SOCIETAL ARTS AND SKILLS			
Counting and Sorting	6.1	7.4	10.3
Imitative Action	4.2	4.3	4.6
Drawing	10.0	13.0	17.0
Number Questions	4.1	4.5	5.6
LANGUAGE COMPETENCE			
Word Knowledge	14.7	16.3	19.8
Verbal Fluency	9.4	10.8	13.4
Opposite Analogies	3.5	3.8	4.8

Note.—Based on partial standardization sample containing a minimum of N=50 at each six month age level (Psychological Corporation, personal communication, 1971).

APPENDIX D

PRE- AND POSTTEST MEAN RATINGS ON THE SCHAEFFER BEHAVIOR INVENTORY ITEMS FOR TEEM CHILDREN

ITEMS:	Pretest		Posttest		Gain	
	Group		Group		Group	
	A N=82	B N=36	A N=82	B N=36	A N=82	B N=36
Task Orientation:	4.5	4.8	4.7	4.8	.2	0
Pays attention to what he's doing when other things are going around him.	4.4	4.7	4.6	4.7	.2	0
Stays with a job until he finishes it.	4.5	4.7	4.8	5.0	.3*	.3
Becomes very absorbed in what he is doing.	4.5	4.8	4.7	4.9	.2*	.1
Works earnestly at his classwork. Doesn't take it lightly.	4.5	4.9	4.5	4.7	0	-.2
Watches carefully when a teacher or classmate is showing how to do something.	4.5	4.9	4.6	4.7	.1	-.2
Extraversion:	4.8	4.9	4.9	5.2	.1	.3
Tries to be with another person or group of people.	4.8	4.9	5.1	5.2	.3*	.3
Likes to take part in activities with others.	5.2	5.3	5.2	5.7	0	.4*
Enjoys being with others.	5.2	5.3	5.2	5.6	0	.3
Seeks social contact with others.	4.9	5.1	4.9	5.3	0	.2
Does not wait for others to approach him, but makes the first friendly move.	4.1	4.1	4.2	4.1	.1	0
Hostility:	2.9	2.9	3.1	2.9	.2	0
Gets impatient or unpleasant if he can't get what he wants when he wants it.	3.3	3.3	3.6	3.1	.3	-.2
Slow to forgive when offended.	3.0	2.8	2.9	2.7	-.1	-.1
Stays angry for a long time after a quarrel.	2.5	2.4	2.7	2.6	.2	.2
Complains or whines if he can't get his own way.	3.0	3.0	3.2	3.0	.2	0
Angry when he has to wait his turn or share with others.	2.8	2.9	3.0	2.8	.2	-.1

Note.—The numbers reflect a verbal Likert-type scale: never, almost never, occasionally, half the time, frequently, almost always, always, with "never" assigned a value of one and "always" a value of seven.

*$p \leq .05$.

APPENDIX E

POSTTEST MEAN RATINGS ON THE SCHAEFFER BEHAVIOR INVENTORY ITEMS FOR TEEM AND COMPARISON GROUP CHILDREN

ITEMS:	TEEM N=104	Comparison N=26
Task Orientation:	4.6	4.5
Pays attention to what he's doing when other things are going around him.	4.5	4.6
Stays with a job until he finishes it.	4.8	4.5
Becomes very absorbed in what he is doing.	4.7	4.7
Works earnestly at his classwork. Doesn't take it lightly.	4.5	4.3
Watches carefully when a teacher or classmate is showing how to do something.	4.6	4.3
Extraversion:	4.8	4.7
Tries to be with another person or group of people.	5.0	5.2
Likes to take part in activities with others.	5.1	4.7
Enjoys being with others.	5.2	5.1
Seeks social contact with others.	4.8	4.7
Does not wait for others to approach him, but makes the first friendly move.	4.8	3.9
Hostility:	3.1	3.4
Gets impatient or unpleasant if he can't get what he wants when he wants it.	3.6	3.4
Slow to forgive when offended.	2.9	3.2
Stays angry for a long time after a quarrel.	2.7	3.2
Complains or whines if he can't get his own way.	3.3	3.6
Angry when he has to wait his turn or share with others.	3.0	3.4

Note.--The numbers reflect a verbal Likert-type scale: never, almost never, occasionally, half the time, frequently, almost always, always, with "never" assigned a value of one and "always" a value of seven.

APPENDIX F

CLASSROOM OBSERVATION PROCEDURE VARIABLE CONSTRUCTS GROUPED BY COGNITIVE AND SOCIO-AFFECTIVE CATEGORIES

Variable Constructs:	Variables:
	Cognitive
Adult instruction	Adult informs children Adult informs children - academic Adult informs children with objects Adult informs children without objects Adult direct questioning of child
Child-initiated learning	Child self-learning Child self-learning with objects Child self-learning without objects Child teaching a different child
	Socio-Affective
Adult attention – five children or less	Adult communication, focus: 1 child Adult communication, focus: 2 children Adult communication, focus: small group Adult praise to child Adult acknowledgement of child
Child self-expression	Child self-expression Child shares life experience Child asking simple question Child asking thought question
Single child-adult interactions	Child response to adult direct question Child response to adult direct question – academic Child response, adult corrective feedback Child response, adult praise Child response, adult praise and corrective feedback Child response, adult acknowledgement Adult asks child a thought question Child positive affect toward adults Adult positive affect toward child Adult to child positive touch
Child-child interactions	Child teaching different child Child interaction with other children
Positive behavior	Child response, adult praise Adult praise to child Adult positive corrective feedback All positive affect Child positive affect toward adults Adult positive affect toward children Adult to child positive touch
Negative behavior	Adult negative behavior Adult negative corrective feedback Child negative behavior

APPENDIX G

FREQUENCIES OF CLASSROOM OBSERVATION PROCEDURE VARIABLE CONSTRUCTS FOR THREE CLASSROOMS

Variable Constructs:	TEEM High	TEEM Low	Comparison
Adult Instruction	431.170	381.384	776.376
Child-Initiated Learning	44.088	26.664	14.472
Adult Attention — Five Children or Less	1180.440	1129.440	1228.704
Child Self-Expression	141.864	135.720	114.000
Single Child-Adult Interactions	250.033	178.344	82.464
Child-Child Interactions	37.512	27.144	8.760
Positive Behavior	109.608	121.848	85.776
Negative Behavior	19.296	2.928	20.448

REFERENCES

Bissell, J.S. The cognitive effects of preschool programs for disadvantaged children. Unpublished doctoral dissertation, Harvard University, 1970.

Campbell, D.T. & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Compensatory education: a national debate. Vol. 3. Disadvantaged Child.* New York: Bruner/Mazel, 1970, pp. 185-207.

Campbell, D.T. & Stanley, J.C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1963.

Cicirelli, V.G. *The impact of Head Start; an evaluation of the effects of Head Start on children's cognitive and affective development.* Washington, D.C.: Office of Economic Opportunity, 1969.

Cronbach, L.J. Course improvement through evaluation. *Teacher's College Record,* 1963, *64,* pp. 672-683.

DiLorenzo, L.T. & Salter, R. An evaluative study of prekindergarten programs for educationally disadvantaged: follow-up and replication. *Exceptional Children,* 1968, *35,* pp. 111-119.

Flanders, N. *Interaction analysis in the classroom.* Ann Arbor, Michigan; School of Education, University of Michigan, 1965.

Goldupp, O. Investigation of independent child behavior in an open classroom setting. Unpublished master's thesis, University of Arizona, 1972.

Grimmett, S., Underwood, B. & Brackney, E. Influence of behavior settings on role of inappropriate and appropriate behavior. Arizona Center for Early Childhood Education, University of Arizona, 1970.

Henderson, R.W. Environmental predictors of academic performance of disadvantaged Mexican-American children. *Journal of Clinical and Consulting Psychology,* 1972, *38,* p. 297.

Hess, R. & Shipman, V. Early experience and the socialization of cognitive modes in children. *Child Development,* 1965, *36* (4), pp. 869-886.

Hughes, M., Wetzel, R.J. & Henderson, R.W. Tucson Early Education Model. Arizona Center for Early Childhood Education, University of Arizona, 1969.

Madow, W. Some comments on the report, "The impact of Head Start." *Brittanica Review of American Education,* July, 1969.

Medley, D.M. & Mitzel, H.E. Measuring classroom behavior by systematic observation. In N.L. Gage (Ed), *Handbook of research on teaching.* Chicago: Rand McNally, 1963, pp. 247-328.

Rentfrow, R.K. Intellectual and language development of disadvantaged preschool children as a function of three preschool curriculum models. Paper presented at the American Educational Research Association Convention, New York, 1971.

Smith, M.S. & Bissell, J.S. Report analysis: The impact of Head Start. *Harvard Educational Review,* 1970, *40,* pp. 51-104.

Weikart, D.P. Relationship of curriculum, teaching, and learning in preschool education. In J.C. Stanley (Ed.), *Preschool programs for the disadvantaged: Five experimental approaches to early childhood education.* Baltimore: John Hopkins Press, 1972, pp. 22-66.

White, R.K. & Lippitt, R. *Autocracy and democracy: An experimental inquiry.* New York: Harper, 1960.